



Università degli Studi di Cagliari

## **DOTTORATO DI RICERCA**

DIRITTO DELL'ATTIVITÀ AMMINISTRATIVA  
INFORMATIZZATA E DELLA COMUNICAZIONE PUBBLICA  
Ciclo XXVIII

### **TITOLO TESI**

LA FUNZIONE CONOSCITIVA DELLE AMMINISTRAZIONI  
PUBBLICHE NELL'ERA DEI BIG DATA  
Settore/i scientifico disciplinari di afferenza  
IUS/10

Presentata da: MATTEO FALCONE  
Coordinatore Dottorato PROF. FELICE ANCORA  
Tutor PROF. ENRICO CARLONI

Esame finale anno accademico 2014 – 2015

# Indice

## LA FUNZIONE CONOSCITIVA DELLE AMMINISTRAZIONI PUBBLICHE NELL'ERA DEI BIG DATA.

<b>Introduzione</b> .....	4
<b>Capitolo 1 – Il fenomeno dei <i>Big data</i>.</b>	
1.1. Che cosa intendiamo per “ <i>Big data</i> ”.....	12
1.2. I caratteri dei <i>Big data</i> e la loro dimensione quantitativa.....	15
1.3. La prima dimensione qualitativa del fenomeno: l'evoluzione tecnologica;.....	20
1.4. L'evoluzione culturale come seconda dimensione quantitativa del fenomeno: dalla logica degli <i>Small data</i> alla logica dei <i>Big data</i> .....	24
1.4.1. La datizzazione della realtà.....	26
1.4.2. Dal “perché” al “che cosa”: osservare (e non spiegare) e prevedere i fenomeni.....	27
1.4.3. L'estrema valorizzazione dei dati e le molteplici finalità del loro utilizzo.....	34
1.5. Le tipologie di utilizzo dei <i>Big data</i> : potenzialità e limiti;.....	37
<b>Capitolo 2 – La dimensione giuridica dei <i>Big data</i>.</b>	
2.1. Il rilievo giuridico dei <i>Big data</i> .....	48
2.2. <i>Big data</i> , mercati e concorrenza.....	50
2.3. <i>Big data</i> e il profilo individuale della tutela dei dati personali.....	55
2.4. I <i>Big data</i> e il profilo collettivo della tutela dei dati individuali.....	62
2.5. <i>Big data</i> , sorveglianza e prevenzione ai fini di sicurezza e controllo.....	66
2.6. <i>Big data</i> , potere e asimmetrie informative.....	71
2.7. I <i>Big data</i> e la regolazione pubblica.....	76
2.8. I <i>Big data</i> e le amministrazioni pubbliche.....	81
<b>Capitolo 3 – I <i>Big data</i> nelle amministrazioni pubbliche</b>	
3.1. Le amministrazioni e i <i>Big data</i> : premessa metodologica.....	83
3.2. L'azione amministrativa e i <i>Big data</i> : un approccio giuridico.....	85
3.2.1. Le procedure di accertamento standardizzato dell'amministrazione finanziaria: un modello per comprendere i <i>Big data</i> nel procedimento amministrativo.....	88
3.2.2. Il valore giuridico degli strumenti accertativi standardizzati e le dinamiche procedurali che innescano: quale lezione per l'elaborazione dei <i>Big data</i> nell'azione amministrativa.....	93

3.3. L'organizzazione amministrativa e i <i>Big data</i> : i primi tentativi di organizzare i <i>Big data</i> .....	99
3.3.1. Le esperienze di organizzazione dei <i>Big data</i> : ISTAT, INPS e INAIL..	101
3.3.1.1. Le politiche di organizzazione dei <i>Big data</i> dell'Istituto nazionale di Statistica (ISTAT).....	104
3.3.1.2. Le politiche di organizzazione dei <i>Big data</i> dell'Istituto nazionale di previdenza sociale (INPS) e dell'Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro (INAIL);.....	109
3.2. L'estrema valorizzazione del patrimonio informativo come fil rouge delle politiche di organizzazione dei <i>Big data</i> ;.....	111
3.4. <i>Big data</i> , amministrazioni pubbliche e cittadino: elementi per una nuova relazione cittadino-amministrazione?.....	113
3.4.1. Il caso dell'epatite C: come l'elaborazione e l'integrazione dei dati da parte dei cittadini influisce sulla decisione pubblica.....	115
3.4.2. L'elaborazione dei <i>Big data</i> da parte dei cittadini come nuovo strumento di equilibrio del potere informativo.....	117

#### **Capitolo 4 – Il *Big data government*: governare i *Big data*, governare con i *Big data*.**

4.1. Le motivazioni di una rinnovata disciplina dell'azione e dell'organizzazione amministrativa.....	120
4.2. Principio di legalità e tipizzazione del potere conoscitivo delle amministrazioni..	122
4.3. Una nuova funzione conoscitiva delle pubbliche amministrazioni.....	124
4.4. <i>Big data</i> e <i>Open data</i> : differenze concettuali e nuovi equilibri tra trasparenza e <i>privacy</i> .....	127
4.5. Una nuova cittadinanza amministrativa con i <i>Big data</i> .....	132

## INTRODUZIONE

L'utilizzo dei dati nel settore pubblico e nel settore privato per le più svariate finalità non è una conseguenza del fenomeno dei *Big data*. Da sempre il settore pubblico acquisisce, archivia, elabora, in una parola, gestisce i dati per legiferare, per governare un territorio, per espletare una funzione e per tutelare gli interessi pubblici a cui quella funzione è preposta, per garantire i diritti dei cittadini. Allo stesso modo, nel settore privato, i dati sono sempre stati utilizzati per accrescere le possibilità di guadagno, aumentare l'offerta dei servizi, migliorare la produttività dell'azienda e organizzarla in maniera più efficiente.

Lo sviluppo tecnologico, in particolare quello informatico, e la diffusione capillare della rete, di Internet, hanno incrementato notevolmente la quantità di dati a disposizione. Oggi è possibile, grazie a questa combinazione di strumenti (la tecnologia e la rete) non solo archiviare ed elaborare velocemente e in quantità notevolmente maggiori i dati, ma è possibile perfino produrre dati da azioni, da cose o da persone, da fenomeni prima impensabili. Si pensi alla diffusione di sensori che, se impiantati all'interno del nostro organismo, riescono a misurare qualsiasi cosa ci accada (da un'infezione in corso all'analisi dei semplici parametri biometrici) o a sensori che riescono a rilevare qualsiasi cambiamento inerente alla struttura portante di un ponte o di un edificio oppure si pensi ancora alla possibilità di registrare migliaia di dati inerenti alle nostre azioni e alle nostre navigazioni su internet.

I dati, quindi, si stanno moltiplicando in maniera esponenziale. A questo cambiamento quantitativo si stanno affiancando, però, dei cambiamenti molto più rilevanti: un cambiamento qualitativo e uno culturale. Stanno cambiando, cioè, sia le modalità di utilizzo dei dati, la complessità delle tecnologie utilizzate, non solo per archiviare questa esplosione di dati, ma soprattutto per elaborarla e per ottenerne risultati senza precedenti, sia l'approccio culturale all'importanza di una raccolta dei dati organizzata e ragionata, di una loro elaborazione meno "guidata" e di una loro utilizzazione e, in particolare, riutilizzazione.

Le nuove tecnologie informatiche, infatti, stanno dando la possibilità di archiviare ed elaborare in maniera del tutto differente e maggiormente efficace, in una parola in maniera qualitativamente maggiore, questa enorme quantità di dati; stanno rendendo obsoleti i precedenti sistemi di archiviazione e di elaborazione diffusi tra i soggetti pubblici e privati; stanno aprendo prospettive di analisi prima impensabili.

Un miglioramento tecnologico che è stato affiancato da (ma ha anche indotto) un cambiamento culturale verso il dato e la sua gestione. I nuovi approcci culturali tendono ad essere sempre più diffusi e sempre più irreversibili.

In primo luogo l'esplosione quantitativa e il salto qualitativo legato alle nuove soluzioni tecnologiche stanno, innanzitutto, modificando la percezione dell'importanza dei dati: possedere e sapere gestire i dati, sapere creare più informazioni possibili dai dati, sta diventando una necessità sempre maggiore per le imprese, per i governi, per le amministrazioni pubbliche. In una società sempre più complessa e sempre più veloce diventa più complicato prevalere sui concorrenti nel mercato, prendere delle decisioni, amministrare e governare con cognizione di causa un territorio. L'informazione, la conoscenza diventa fondamentale per raggiungere i propri obiettivi (di leadership nel settore o di governo) e diventa fondamentale “conoscere” continuamente il contesto in cui si agisce. Il proverbiale “la conoscenza è potere”, che sia inteso come potere di battere la concorrenza o come carattere del potere pubblico trova in questi fenomeni sempre maggiore legittimazione.

In secondo luogo cominciano a cambiare anche le preferenze sui metodi di rilevazione dei dati e di interpretazione dei loro risultati. Dall'interpretazione dei risultati basata sul rapporto causa-effetto si comincia ad intravedere la tendenza a volere interpretare i risultati delle elaborazioni dei dati attraverso le correlazioni: dalla volontà/capacità di comprendere i motivi di un fenomeno, del “*perché*” di un fenomeno, si sta passando alla volontà/capacità di comprendere anche “*che cosa*” sta accadendo e a farlo velocemente. Con le correlazioni di enormi quantità dei dati, secondo alcuni studiosi, come vedremo nelle pagine che seguono, si raggiungerebbe una capacità di elaborazione che potrebbe essere quasi predittiva di un fenomeno.

In terzo luogo alle due tendenze precedenti si affianca una tendenza culturale alla “datizzazione” della realtà. La tendenza, cioè, a trasformare in dato qualsiasi aspetto della nostra esistenza: le persone, le cose, i sentimenti, i fenomeni ambientali, sociali, economici, politici.

Infine, il cambiamento qualitativo sta producendo un cambiamento culturale anche in merito alle finalità secondarie dell'utilizzo dei dati, molte volte lontanissime dalle finalità iniziali, da quelle motivazioni, cioè, che avevano portato i soggetti pubblici e privati a raccogliere determinati dati e a produrre quelle determinate informazioni.

Il tema dei *Big data* è caratterizzato senza dubbio dall'elemento quantitativo, dalla quantità dei dati che è possibile reperire o produrre, ma sprigiona il suo potenziale dalla combinazione della *quantità* con la *qualità*, con questi tre grandi fenomeni conoscitivi che abbiamo appena accennato: la correlazione/quasi predizione, la datizzazione della

realtà che ci circonda, la moltiplicazione del valore dei dati attraverso un loro ipotetico continuo riutilizzo. Questi fenomeni conoscitivi costituiscono una sfida di non poco momento certamente per il settore privato, ma in particolar modo per il settore pubblico.

Lo dimostra la crescente rilevanza che gli stessi soggetti pubblici e privati danno al fenomeno. Per portare qualche elemento fattuale è necessario segnalare come il *Data Warehouse Institute* (TWDI), una società che si occupa di business intelligence, ha rilevato come già nel 2011 un terzo delle aziende statunitensi usasse i *Big data*<sup>1</sup>. Per quanto riguarda l'Italia vale la pena riportare sommariamente i risultati di un recente studio che ha dimostrato come, nel settore privato, l'attenzione e la consapevolezza dell'importanza del fenomeno stia crescendo notevolmente tra le imprese medie e medio-grandi<sup>2</sup>: il 9% delle imprese rientranti nel campione sta già attuando o sta per attuare una strategia per affrontare il fenomeno dei *Big data* mentre un nutritissimo 57% delle imprese del campione è in una fase di studio del fenomeno e dei suoi possibili vantaggi. Possiamo dire, quindi, che il 66% del campione percepisce l'importanza del fenomeno dei *Big data* e ha agito o sta agendo per sfruttarne le potenzialità. In realtà la percezione che il fenomeno sia rilevante e che possa riguardare la propria attività è, in percentuale molto maggiore. Il 24% delle imprese del campione comprende, infatti, il valore e i benefici che i *Big data* potrebbero portare alla loro attività, ma hanno deciso di dare precedenza ad investimenti in altre tecnologie ICT. Quindi, in definitiva, il 90% delle imprese, che ricordiamo sono imprese medie e medio-grandi italiane, ha la percezione dell'importanza dei *Big data* per la loro attività e dei vantaggi che potrebbero derivare da un loro più costante utilizzo.

Nel settore pubblico, allo stesso modo, la percezione internazionale della rilevanza del fenomeno sta crescendo costantemente. Nella politica, soprattutto negli Stati Uniti, i *Big data* sono stati utilizzati frequentemente per organizzare la propaganda a fini elettorali: le campagne di Obama sono state un esempio emblematico. Nelle amministrazioni pubbliche, che è il profilo che interessa maggiormente la nostra analisi, allo stesso modo, i *Big data* si stanno diffondendo soprattutto nei settori in cui i dati sono maggiormente utilizzati, o meglio maggiormente utili ai fini dell'azione amministrativa - come il settore della conoscenza pubblica, della sanità, della difesa e della sicurezza interna, della tutela ambientale - ma in generale e potenzialmente riguarda tutte le amministrazioni in tutti i settori.

Basta consultare, infatti, gli ultimi documenti della Casa Bianca, in particolare il *Fact Sheet: Big Data Across the Federal Government*, del marzo 2012 prodotto dall'*Office of*

---

1 TWDI, *Big Data Analytics*, 14 settembre 2011. Sul sito è possibile anche consultare una serie, anche recentissimi, di report sul tema.

2 P. PASINI – A. PEREGO (A CURA DI), *Big data: nuove fonti di conoscenza aziendale e nuovi modelli di management*, Rapporto di ricerca per IBM, SDA Bocconi – School of Management, 2012.

*Science and Technology Policy* dell'*Executive Office of the President*. In questo documento gli uffici presidenziali mettono in fila tutta una serie di attività in cui l'amministrazione federale statunitense sta utilizzando i *Big data*. Questo breve documento, su cui torneremo nelle pagine seguenti, è determinante, da un lato, per rilevare come, in particolare negli Stati Uniti, alcune amministrazioni pubbliche stanno cominciando ad analizzare il fenomeno dei *Big data* e stanno cominciando a prendere atto di come esso possa modificare il loro modo di utilizzare i dati e, conseguentemente, amministrare la cosa pubblica, dall'altro, per osservare come il documento della Casa Bianca costituisca, se non un *unicum*, una rarità: le amministrazioni pubbliche in generale, comprese quelle italiane, stentano ancora a considerare la prospettiva dei *Big data* come una prospettiva concreta di evoluzione delle modalità di utilizzo dei dati, anche se ci sono delle importanti eccezioni.

Che i *Big data* siano una sfida importate per il settore privato e per il settore pubblico lo dimostra anche la crescente attenzione che gli studiosi delle varie discipline stanno dando al fenomeno. Se gli studiosi di informatica e matematica hanno sottolineato la fattibilità tecnica e l'esistenza di tecnologie che oggi possono utilizzare i *Big data*, la dottrina giuridica, politologico-sociologica ed economica ha incominciato ad indagare i vantaggi, i limiti e i rischi che l'utilizzo dei *Big data* comporta. Come sarà possibile leggere nelle pagine seguenti, i profili di analisi prevalenti tra gli studiosi che si sono occupati di questa materia sono legati alla fattibilità tecnica delle tecnologie che elaborano i *Big data*, al rapporto tra costi della tecnologia e vantaggi economici, alle prospettive di sviluppo che l'utilizzo massivo dei dati può apportare alle imprese e alla società in generale, al rapporto tra *Big data* e privacy.

La dottrina giuridica, in particolare, ha trattato prevalentemente quest'ultimo aspetto, sottolineando come i *Big data* e i fenomeni conoscitivi ad esso sottesi (come la "datizzazione" e il metodo della correlazione) comportino dei rischi per la tutela dei dati personali.

In realtà i profili appena menzionati, che attengono alla tutela del singolo e delle sue situazioni giuridiche, non sembrano essere i soli profili giuridici inerenti ai *Big data* degni di essere analizzati e approfonditi. Anche se sono aspetti quasi per nulla indagati dalla dottrina giuridica, in particolar modo da quella amministrativistica, i fenomeni conoscitivi prodotti dall'utilizzo massivo dei dati hanno un impatto rilevante non solo sulla *privacy* dei cittadini, ma anche su processi decisionali in cui essi vengono utilizzati (sia da soggetti pubblici, sia da soggetti privati), sull'organizzazione e sull'azione delle amministrazioni pubbliche.

Con questo scritto si tenterà di descrivere il fenomeno dei *Big data*, di sottolineare la sua attualità e la profonda influenza che sta avendo sull'agire democratico e

amministrativo dei governi e delle pubbliche amministrazioni, senza tralasciare il suo impatto sui regimi e sui sistemi di tutela dei diritti per i cittadini. La prospettiva sarà, ovviamente, una prospettiva giuridica - che avrà una particolare attenzione all'approccio al fenomeno avuto dalle altre discipline, in particolare economiche, filosofiche e politico-sociali – e che riporrà una maggiore attenzione all'impatto che i *Big data* stanno avendo e avranno potenzialmente sulle amministrazioni pubbliche

Con questi intenti, si cercherà, nel primo capitolo, di analizzare il fenomeno nella sua interezza sottolineando quali sono le caratteristiche, le innovazioni tecnologiche e culturali che stanno apportando, i modi in cui sono stati utilizzati fattivamente fino a questo momento, le potenzialità e limiti dei *Big data*.

Si cercherà, nel secondo capitolo, proprio partendo dalle tipologie di utilizzo dei *Big data*, di comprendere quali sono le implicazioni giuridiche che i *Big data* producono nel nostro ordinamento prendendo in considerazione le prime analisi e i primi studi che si sono occupate della questione. Vedremo come l'elaborazione dei *Big data* sta modificando le dinamiche concorrenziali in molti mercati, come quello dei servizi e quello digitale, e conseguentemente sta mettendo in crisi aspetti fondamentali della disciplina *anti-trust*. L'elaborazione e l'utilizzo dei *Big data*, inoltre, sta scardinando tutte le discipline a tutela dei soggetti deboli nel rapporto tra privati, come la disciplina a tutela del consumatore e quella a protezione dei dati personali: i *Big data* stanno paralizzando i principi fondamentali su cui si fondavano entrambe le discipline e le analisi prese in considerazione stanno prendendo atto di come entrambe le discipline debbano convergere su rinnovati parametri e strumenti di tutela.

Un altro aspetto rilevante del fenomeno dei *Big data*, che verrà sottolineato nel secondo capitolo, riguarda la sua capacità di consegnare nelle mani di coloro che possiedono questo tipo di dati e di informazioni e le tecnologie giuste per utilizzarle, che siano essi soggetti pubblici o privati, un potere di controllo sociale rilevante: si pensi agli effetti che la categorizzazione e la profilazione di massa producono non solo sulla *privacy* dei cittadini, ma anche sulle loro libertà fondamentali.

Le implicazioni giuridiche rilevanti non sono solo, però, quelle che scaturiscono da un approccio analitico “difensivo” al fenomeno dei *Big data*, un approccio che vede i *Big data* come qualcosa da cui difendersi, ma possono scaturire anche da un approccio analitico differente, che metta al centro dell'analisi come si possano utilizzare i *Big data*, e che tocchi aspetti importanti del diritto pubblico e del diritto amministrativo. Implicazioni che attengono a come i processi decisionali pubblici vengono influenzati dall'elaborazione dei *Big data* e a come le opzioni regolatorie e le scelte regolative degli organi politici possono essere modificati, al rapporto tra esecutivo e legislativo, al rapporto tra soggetti privati e soggetti pubblici all'interno dei processi decisionali.

Implicazione che, come vedremo, sono tutte legate al *fil rouge* della capacità dei *Big data* di produrre delle forti asimmetrie informative tra coloro che possiedono i dati e le informazioni e coloro che non possiedono né i primi e né i secondi.

Dopo avere elencato quali sono le implicazioni giuridiche più rilevanti e avere sottolineato come l'azione e l'organizzazione amministrativa non sono esenti dagli effetti del fenomeno, nel terzo capitolo, si cercherà di comprendere come le pubbliche amministrazioni stanno utilizzando concretamente i *Big data*, al fine di costruire una base concreta da cui partire per comprendere quali implicazioni un utilizzo generalizzato dell'elaborazione dei *Big data* nell'azione e nell'organizzazione amministrativa, intesa come organizzazione del patrimonio informativo pubblico, possano sorgere. L'analisi delle dinamiche procedurali e delle tutele per i contribuenti all'interno dei procedimenti tributari con accertamento standardizzato, che utilizzano strumenti che funzionano attraverso funzioni matematiche e statistiche come gli studi di settore e il redditometro, da un lato, e i primi tentativi degli Istituti nazionali più importanti (ISTAT, INPS e INAIL) di dotarsi di strumenti tecnologici di raccolta e conservazione dei *Big data*, di organizzarli e cominciare ad elaborarli per le rispettive attività istituzionali, dall'altro, ci offrono l'opportunità di rilevare concretamente quali sono le necessità ordinarie e le problematiche giuridiche che l'elaborazione e l'utilizzo dei *Big data* possono fare sorgere all'interno delle amministrazioni pubbliche.

Alla luce di queste considerazioni, nel quarto e ultimo capitolo si cercherà di tirare le fila della ricostruzione e di sottolineare come l'elaborazione e l'utilizzo dei *Big data*, le sue potenzialità conoscitive, la sua capacità di portare nuove conoscenze sui fenomeni, le sue capacità predittive modificano profondamente l'attività conoscitiva delle amministrazioni pubbliche, la quale diviene sempre più una funzione a se stante.

La dimensione dell'azione e quella dell'organizzazione della funzione conoscitiva delle pubbliche amministrazioni, con i *Big data*, diventano delle chiavi di lettura sempre più calzanti per spiegare le prospettive evolutive che i *Big data* aprono per le amministrazioni pubbliche.

I *Big data* modificano, vedremo quanto in positivo o in negativo, il principio di informalità e il metodo inquisitorio che caratterizzano l'attività istruttoria nel nostro ordinamento e mettono in evidenza come l'attività conoscitiva delle pubbliche amministrazioni non possa essere più considerata come un'attività strumentale all'azione amministrativa, a costo di svilirne la sua effettività di funzione atta a permettere all'amministrazione di conoscere i bisogni e le necessità delle comunità di riferimento e sulle quali le proprie decisioni producono effetti rilevanti. La stessa decisione amministrativa, proprio grazie alle potenzialità conoscitive che l'elaborazione dei *Big data* produce, assume dei caratteri differenti, di maggiore certezza, di maggiore

completezza. Questo non può esimerci dal ritornare ad affrontare il rapporto, sempre complesso, tra verità materiale, conoscenza oggettiva dei fenomeni e decisione amministrativa. Il concreto utilizzo dei *Big data* all'interno del procedimento amministrativo, l'intensità del loro utilizzo e l'esclusività dei risultati della loro elaborazione come fondamento della decisione amministrativa, inoltre, può modificare la discrezionalità amministrativa, ma anche fare riflettere sulla trasformazione dell'attività conoscitiva delle amministrazioni in un vero e proprio potere conoscitivo, capace di comprimere i diritti dei cittadini.

La possibilità di elaborare e utilizzare i *Big data* non modifica solo la dimensione dell'azione della funzione conoscitiva, ma ne modifica anche la dimensione organizzativa, la gestione del patrimonio informativo pubblico. Un patrimonio informativo che con i *Big data* diventa più complesso, più eterogeneo e che pone all'ordinamento la necessità di adeguare a questa nuova eterogeneità la garanzia di qualità dei dati pubblici e il regime dei dati in generale, il coordinamento amministrativo tra i vari livelli di governo.

Il regime dei dati, della loro gestione e della loro utilizzazione e riutilizzo da parte dei cittadini deve tenere conto delle necessità legate al rispetto della qualità dei dati e del possibile affidamento che i cittadini ripongono sulla provenienza pubblica dei dati, della *privacy* dei cittadini e delle nuovi potenziali rischi per la riservatezza dei dati personali che l'utilizzo dei *Big data* si porta dietro, ma deve anche garantire la massima diffusione dei dati e delle informazioni. Politiche di *Open data* sui *Big data* diventano in un contesto di forte accentramento informativo le uniche politiche pubbliche capaci di ristabilire gli equilibri democratici che le asimmetrie informative create dai *Big data* tendono ad incrinare. La diffusione e il riutilizzo dei *Big data* pubblici crea le condizioni per creare una nuova cittadinanza consapevole e capace di raccogliere e formare le informazioni attraverso le quali controllare l'operato delle amministrazioni e di partecipare consapevolmente ed attivamente alla decisione pubblica, garantendo anche un ritorno informativo per le amministrazioni, le quali magari non possedevano determinate informazioni.

Il fenomeno dei *Big data* si dimostra, quindi, un fenomeno molto complesso, ricco di implicazioni giuridiche e che costituisce un fenomeno da cui gli studi amministrativistici non possono più prescindere, non solo quando si trattano i temi legati alla gestione dei dati in possesso delle amministrazioni, ma quando si trattano i temi più rilevanti riguardanti l'azione e l'organizzazione amministrativa e riguardanti i rapporti tra amministrazione e cittadini.

I *Big data* sono un fenomeno che travolge totalmente le amministrazioni pubbliche e il diritto amministrativo, che con questo lavoro si vuole cominciare ad analizzare, con un

approccio laico, senza pregiudizi, con l'unico intento di capirlo e di sottolinearne l'importanza prospettica.

## Capitolo 1

### Il fenomeno dei *Big data*

Sommario: 1. Che cosa intendiamo per “*Big data*”; 2. I caratteri dei *Big data* e la loro dimensione quantitativa; 3. La prima dimensione qualitativa del fenomeno: l'evoluzione tecnologica; 4. L'evoluzione culturale come seconda dimensione qualitativa del fenomeno: dalla logica degli *Small data* alla logica dei *Big data*.; 4.1. La datizzazione della realtà.; 4.2. Dal “perché” al “che cosa”: osservare (e non spiegare) e prevedere i fenomeni.; 4.3. L'estrema valorizzazione dei dati e le molteplici finalità del loro utilizzo.; 5. Le tipologie di utilizzo dei *Big data*: potenzialità e limiti.

#### 1. *Che cosa intendiamo per “Big data”.*

Prima di cercare di comprendere il rilievo giuridico del fenomeno dei *Big data* e il suo impatto sul settore pubblico e in particolare sull'azione e sull'organizzazione delle pubbliche amministrazioni è utile cercare di delineare che cosa si intende descrivere con il termine *Big data*, quali sono le sue caratteristiche, le sue potenzialità e i suoi limiti.

Anche se il termine *Big data* comincia a circolare negli ambienti informatici intorno al 2010, il fenomeno dell'esplosione dei dati nasce nella genomica e nell'astronomia già negli anni 2000<sup>3</sup> e alcune caratteristiche peculiari dei *Big data* vengono già individuate, anche se in maniera imperfetta e neanche legate al termine in analisi, nel 2001. In un *report*, oramai molto citato, della allora META Group (oggi Gartner Inc.)<sup>4</sup>, Douglas Laney, famoso analista dei dati che lavora presso la stessa società, già cominciava a sottolineare come stesse crescendo la mole di dati a disposizione, la velocità con cui fosse possibile produrre i dati e l'eterogeneità dei loro formati e delle loro fonti di produzione<sup>5</sup>. Considerazioni che, come vedremo quando si parlerà delle caratteristiche dei *Big data*, sono rimaste fondamentali per spiegare e definire il fenomeno, e che, tenendo presente la loro datazione (6 febbraio 2001), sembrano essere state davvero profetiche.

Occorre chiarire fin da ora che c'è ancora un acceso dibattito sulla definizione di *Big data*: ci sono definizioni che tentano di delineare il fenomeno in maniera complessiva, esaltando una o più caratteristiche dei *Big data*; definizioni che mettono in relazione i

---

3 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, trad. it. R. Merlini, Garzanti, Milano, 2013, p. 15.

4 La META Group, oggi Gartner Inc., è una delle maggiori società di consulenza strategica, di ricerca e di analisi dei dati nel settore delle ICT.

5 D. LANEY, *3d data management: Controlling data volume, velocity and variety*, Gartner, Stamford, CT, USA, White Paper, 2001.

*Big data* con gli strumenti tradizionali di conservazione ed elaborazione dei dati, per sottolineare la capacità dei primi di rendere obsoleti i secondi, quindi senza definire davvero che cosa si intende quando si parla di *Big data*; definizioni, infine, che associano il termine *Big data* alle nuove tecnologie che gestiscono questi dati o ai processi che esse innescano<sup>6</sup>.

L'idea iniziale che sottendeva al termine *Big data* era, in realtà, legata soltanto alle tecnologie di lavorazione dei dati molto avanzate che incominciavano ad essere prodotte dalle maggiori aziende informatiche del mondo (come MapReduce di Google o Hadoop di Yahoo!), le quali permettevano (e permettono ancora, in maniera sempre più raffinata) di elaborare un numero molto maggiore di dati<sup>7</sup>. Oggi la maggior parte delle definizioni, quindi, legano ancora il termine *Big data* alle sole tecnologie di elaborazione<sup>8</sup>. La Microsoft, per esempio, definisce i *Big data* come il processo di elaborazione di una enorme e complessa quantità di dati: *Big data* è «the term increasingly used to describe the process of applying serious computing power - the latest in machine learning and artificial intelligence - to seriously massive and often highly complex sets of information»<sup>9</sup>. Il gruppo di lavoro che fa capo al progetto open-source *Method for an Integrated Knowledge Environment* (MIKE), allo stesso modo, definisce i *Big data* come «the high degree of permutations and interactions within a data set», mettendo in rilievo il processo di interazione tra i dati dentro i *dataset* e, quindi, prestando una particolare attenzione ai processi di elaborazione dei dati<sup>10</sup>.

Altre definizioni, invece, prediligono delineare il fenomeno descrivendone gli effetti sulle tecnologie tradizionali, sottolineando il valore aggiunto che questi dati possono apportare alle classiche analisi dei dati. La ORACOL, infatti, definisce i *Big data* come «the derivation of value from traditional relational database-driven business decision making, augmented with new sources of unstructured data», cioè come il valore prodotto dai tradizionali database relazionali accresciuto da un insieme di dati non strutturati<sup>11</sup>. Sempre all'interno di questa tipologia di definizioni può essere inserita quella del *National Institute of Standards and Technology* (NIST), che definisce i *Big data* come i dati che «exceed(s) the capacity or capability of current or conventional methods and systems»<sup>12</sup>.

---

6 Con una simile classificazione anche H. HAN – W. YONGGANG – C. TAT-SENG – L. XUELONG, *Towards scalable systems for big data analytics: a technology tutorial*, *IEEE Access*, 2014, p. 654.

7 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit..

8 A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, Maggioli, Bologna, 2013, p. 19.

9 J.S. WARD – A. BARKER, *Undefined by data: a survey of a Big data definitions*, University of St. Andrews, UK, 20 september 2013.

10 AA.VV., *The Big Data Conundrum: How to Define It?*, MIT Technology Review, 3 october 2013.

11 J.S. WARD – A. BARKER, *Undefined by data: a survey of a Big data definitions*, cit..

12 AA.VV., *The Big Data Conundrum: How to Define It?*, cit..

Nel tempo, però, sono maturate delle definizioni molto più articolate di *Big data* che cercano di descrivere il fenomeno attraverso un'elencazione delle sue caratteristiche principali. Anche in questa occasione Douglas Laney ha aperto la strada ad una definizione di *Big data* oggettivamente più completa, riprendendo le sue considerazioni del 2001. Nel 2012 Laney definisce i *Big data* come «high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization»<sup>13</sup>. Molto più generica, ma collocabile nella stessa categoria definitoria, è quella di Mayer-Schonberger e Cukier, i quali usano l'espressione “*Big data*” per «designare delle cose che si possono fare solo su larga scala, per estrapolare nuove indicazioni o creare nuove forme di valore con modalità che vengono a modificare i mercati, le organizzazioni, le relazioni tra cittadini e governi e altro ancora»<sup>14</sup>. Molto ispirata alla definizione di Laney, invece, quella di De Mauro-Greco-Grimaldi, in cui i *Big data* «represents the information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value»<sup>15</sup>.

La definizione di *Big data*, quindi, resta abbastanza variegata: da *Big data* come nuova tecnologia di gestione dei dati a *Big data* come insieme di dati con determinate caratteristiche che, se elaborati, possono produrre degli effetti inediti sia sulle tradizionali tecnologie, sia sulla società generalmente intesa. La definizione che, però, sembra essere maggiormente convincente e molto più utile per i nostri fini, resta quella di Laney del 2012. In quella definizione, infatti, non solo i *Big data* vengono definiti come un enorme insieme di dati (high volume), eterogeneo nelle fonti e nelle forme (high variety), alimentato continuamente ed in tempo reale (high velocity) - una definizione, quindi, che cerca di chiarire la natura dei *Big data*, al di là della tecnologia preposta alla loro elaborazione – ma vengono descritti anche gli effetti che l'utilizzo dei *Big data* ha sulle decisioni, sulla conoscenza, sui processi. I *Big data*, secondo la definizione di Laney, richiedono tecnologie di elaborazione e di gestione differenti, migliori rispetto alle tecnologie tradizionali, se si vogliono utilizzare per rafforzare i processi decisionali, per migliorare le conoscenze interne e, conseguentemente, per ottimizzare i processi organizzativi. Laney definisce i *Big data* ad uso e consumo delle società private, che vogliono migliorare la produttività della propria azienda, ma la sua definizione è molto utile anche per comprendere quali potrebbero essere gli effetti dei *Big data* su qualsiasi tipologia di attività e di organizzazione, anche pubblica. Se decliniamo, infatti, quella definizione nel settore pubblico potremmo dire che i *Big*

13 M. A. BEYER - D. LANEY, *The importance of big data: A definition*, Stamford, Gartner Retrieved, 21 June 2012.

14 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit..

15 A. DE MAURO – M. GRECO – M. GRIMALDI, *What is big data? A consensual definition and a review of key research topics*, AIP Conference Proceedings, p. 97–104.

*data sono un enorme insieme di dati, eterogeneo nelle fonti di produzione e nelle forme, alimentato continuamente ed in tempo reale, che richiede nuove tecnologie di gestione e il suo utilizzo modifica i processi decisionali, la formazione della conoscenza pubblica e l'azione amministrativa. L'elaborazione e la gestione dei Big data, infatti, possono influire sul processo decisionale democratico, possono accrescere le conoscenze del legislatore e dell'amministrazione sulla realtà che sono preposti a regolare e ad amministrare e possono migliorare l'efficienza e l'efficacia dell'organizzazione e dell'azione legislativa e amministrativa.*

L'obiettivo di questo scritto è proprio descrivere come questo sta avvenendo e quali sono le sue possibili implicazioni giuridiche.

## *2. I caratteri dei Big data e la loro dimensione quantitativa.*

Per comprendere il fenomeno dei *Big data* e per capire fino in fondo i suoi effetti e le sue implicazioni sulla società e sui soggetti pubblici e privati, non basta cercare di ricostruire una sua definizione completa. È fondamentale anche descrivere nella maniera più dettagliata e comprensibile possibile quali sono le caratteristiche dei *Big data* e concentrarci, in prima istanza, su che tipo di dati sono. Cerchiamo di chiarire, cioè, che cosa intendiamo quando parliamo di “volume, velocità, varietà”, concetti che sono ritornati molto spesso quando abbiamo elencato le tipologie di definizioni utilizzate fino ad oggi.

Tutti coloro che studiano il fenomeno dei *Big data*, infatti, sottolineano come i caratteri principali sono riconducibili alle 3V (Volume, Velocità, Varietà)<sup>16</sup>. Questa caratterizzazione dei *Big data*, però, è una caratterizzazione descrittiva, che non assolutizza la presenza costante dei tre caratteri in tutti i casi in cui si parlerà di *Big data*: la maggior parte delle volte, o quantomeno per quanto riguarda le tipologie di *Big data* più diffuse, tutti e tre i caratteri sono presenti, ma in alcuni casi, anche rilevanti, possono essere presenti anche solo due dei tre caratteri in esame (di solito Volume e Velocità).

La prima caratteristica dei *Big data* è certamente quella di essere un insieme di dati enorme, talmente grande da rendere obsolete le tradizionali tecnologie di conservazione dei dati, le quali non riescono più ad archivarle in maniera tradizionale. Una dimensione che nel tempo è destinata soltanto a crescere e che diventerà sempre più rilevante.

---

<sup>16</sup> Per tutti si veda M. A. BEYER - D. LANEY, *The importance of big data: A definition*, Stamford, Gartner Retrieved, 21 June 2012;

Se nel paragrafo precedente abbiamo fatto riferimento in termini generali alla genomica e all'astronomia come ai due settori scientifici in cui l'esplosione di dati è stata repentina e dove, di fatto, è nato embrionalmente il concetto di *Big data*, ora è utile precisare quelle affermazioni soprattutto perché ci paiono emblematiche per esemplificare il concetto di "Volume" e per sottolineare la futura esplosione della quantità di dati in circolazione. Nel settore astronomico, infatti, se nel 2000 erano stati raccolti più dati di quanti ne fossero stati raccolti nell'intera storia dell'astronomia, grazie all'utilizzo di un nuovo telescopio all'interno del programma Sloan Digital Sky Survey (nel New Mexico) e se nel 2010 l'archivio del progetto conteneva 140 terabyte<sup>17</sup> di dati, nel 2016 verrà installato in Cile un nuovo telescopio, il Large Synoptic Survey Telescope, che produrrà la stessa quantità di dati in un solo giorno<sup>18</sup>. Nel settore genomico, allo stesso modo, se gli scienziati hanno messo in sequenza circa i tre miliardi di coppie di genoma umano in circa dieci anni fino al 2003, dal 2013, grazie allo sviluppo della tecnologia, un singolo centro di ricerca è capace di sequenziare la stessa quantità di DNA in un solo giorno<sup>19</sup>.

Un altro settore in cui è possibile rilevare un'esplosione dei dati è quello finanziario. Ogni giorno nella Borsa di New York circolano circa sette miliardi di azioni e i due terzi vengono negoziati attraverso algoritmi matematici che elaborano altrettanti miliardi di dati per prevedere guadagni e rischi<sup>20</sup>.

La percezione della dimensione quantitativa del fenomeno, però, ci diviene ancora più evidente se ricorriamo alle capacità delle aziende della Silicon Valley di reperire e produrre dati da Internet. Google processa oltre 24 petabyte<sup>21</sup> al giorno, un volume pari a mille volte la quantità di tutto il materiale a stampa della Biblioteca del Congresso degli Stati Uniti<sup>22</sup>. Su Facebook si caricano circa dieci milioni di fotografie ogni ora e i suoi iscritti lasciano commenti o "Mi piace" quasi tre miliardi di volte al giorno<sup>23</sup>. Il numero dei messaggi su Twitter cresce del 200% all'anno e nel 2012 ha superato 400 milioni di tweet al giorno<sup>24</sup>.

---

17 È opportuno ricordare che l'unità di misura dei dati in informatica è il byte. Un terabyte corrisponde ad un bilione di byte.

18 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit..

19 K. CUKIER, *Data, data, everywhere*, numero speciale di «The Economist», 27 febbraio 2010;

20 M. HILBERT – P. LOPEZ, *The world's technological capacity to communicate, store and compute information?*, in *International Journal of Communication*, 2012, p. 1042-1055.

21 Tenendo a mente che l'unità di misura della quantità di dati è il byte, un petabyte corrisponde ad un biliardo di byte (un multiplo del byte successivo al terabyte utilizzato in precedenza).

22 T. H. DAVENPORT – P. BARTH – R. BEAN, *How Big data is different*, in *Sloan review*, 30 luglio 2012, p. 43-46.

23 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit.

24 T. GERON, *Twitter's dick costolo: Twitter mobile and revenue beats desktop on same days*, in *Forbes*, 6 giugno 2012.

Qualche studioso ha tentato anche di quantificare il volume dei dati circolanti nel mondo, intesi come qualsiasi dato analogico o digitale che è stato prodotto, archiviato e comunicato. Se nel 2007 i dati erano 300 exabyte<sup>25</sup> di cui solo il 7% era in formato analogico, nel 2013 la quantità di dati mondiali era cresciuta talmente tanto da arrivare a 1200 exabyte totali con una percentuale di dati analogici scesa ad un 2% del totale: un numero triplicato in soli 6 anni<sup>26</sup>. Altri studi segnalano come la quantità di dati archiviati nel mondo cresca quattro volte più in fretta dell'economia mondiale e che elaboriamo soltanto il cinque per mille dei dati che produciamo<sup>27</sup>.

I *Big data* hanno, quindi, come prima caratteristica una dimensione quantitativa enorme, una dimensione che - grazie ai continui sviluppi tecnologici di produzione, conservazione ed elaborazione dei dati - è destinata a crescere in maniera sempre più repentina.

Oltre al volume, i *Big data* hanno la caratteristica di essere anche dati eterogenei non solo nel formato in cui si presentano, ma anche rispetto alle fonti che li producono.

È opportuno, però, fare una piccola premessa prendendo in considerazione alcune nozioni e classificazioni dell'informatica per comprendere la particolarità dei *Big data*, soprattutto sul versante dell'eterogeneità. I dati, nel settore dell'informatica, vengono classificati in base alla possibilità/capacità dei *database* relazionali di archivarli o meno all'interno delle loro tabelle<sup>28</sup>. I dati, quindi, vengono classificati come dati strutturati, se possono essere archiviati ed elaborati utilizzando le banche dati tradizionali, e come dati non strutturati, invece, qualora avessero un formato tale da impedire ai database tradizionali di archivarli ed elaborarli in maniera ordinaria<sup>29</sup>. Quando parliamo di

25 Un exabyte corrisponde ad un trilione di byte ed è un multiplo del byte superiore sia al terabyte che al petabyte.

26 M. HILBERT – P. LOPEZ, *The world's technological capacity to communicate, store and compute information?*, cit..

27 J. GANZ – D. REINSEL, *Big data, bigger digital shadows and biggest growth in the Far East*, 1 dicembre 2012. Il report è stato commissionato dalla EMC Corporation, una delle maggiori aziende internazionali che sviluppa, fornisce e supporta infrastrutture per l'information technology.

28 I database sono un insieme di oggetti: tabelle, indici e codici SQL. Le tabelle contengono i dati; gli indici agevolano la ricerca di un determinato dato; i codici SQL permettono di estendere le funzionalità dei database inserendo nuove funzioni (come la possibilità di elaborare dei risultati dalla serie storica dei dati a disposizione). Tecnicamente le tabelle vengono chiamate “relazioni”. Per database relazionale intendiamo un sistema di archiviazione dei dati che utilizza delle tabelle (ecco perché relazionale) e le mette in relazione per un qualsiasi obiettivo o finalità. I database relazionali sono gli strumenti tradizionali di archiviazione ed elaborazione dei dati più diffusi. Per approfondire il tema dal punto di vista tecnico si v. A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit..

29 L'archiviazione ordinaria di un insieme di dati in un database relazionale avviene attraverso la loro normalizzazione, cioè attraverso un processo di ordinazione dei dati nel database. La normalizzazione avviene con tre operazioni. La prima è l'indicizzazione del dato, cioè creazioni di indici per classificare i dati contenuti nelle tabelle. Gli indici sono strutture di dati ordinate che agevolano il recupero dei dati dalle tabelle quando si ricorre a delle query (cioè quando si interrogano i database). La seconda operazione è la compressione del dato, cioè comprimere lo spazio occupato dal dato, il

eterogeneità dei *Big data*, quindi, ci riferiamo in prima istanza al fatto che essi sono classificabili come dati non strutturati, cioè dati che non hanno un formato standard o hanno una struttura complessa che non permette loro di essere classificati nelle strutture tabellari dei database relazionali tradizionali.

L'eterogeneità, però, dipende anche dal fatto che questa tipologia di dati non strutturati proviene anche da fonti molto differenti tra di loro: sensori, rilevatori di vario genere e composti con diverse tecnologie, ma anche persone che condividono foto, video, commenti sul web<sup>30</sup>. Eterogeneità, dunque, sia come differente struttura del dato sia come differenziazione di fonti di produzione.

L'ultima caratteristica determinante per considerare un insieme di dati come *Big data* è la velocità con cui è possibile reperirli. I sensori, il web, qualsiasi tipologia di tecnologia di riconoscimento e di rilevazione hanno come caratteristica intrinseca quella di produrre in tempo reale una quantità enorme di dati, che rendono ancora più complicato per i database relazionali tradizionali l'archiviazione dei dati e la loro elaborazione: se, per esempio, si installasse un sistema di controllo in un impianto industriale composto da mille sensori e che questi facessero una rilevazione ogni secondo, riceveremmo dati per tutto il giorno fino a raggiungere una quantità giornaliera di 86.400.000 dati al giorno<sup>31</sup>.

Queste tre caratteristiche, quindi, costituiscono la “specialità” dei *Big data* rispetto ai dati tradizionali, in particolar modo caratterizzano la dimensione quantitativa del fenomeno. Come vedremo anche in seguito, il vero discrimine è l'esplosione della quantità dei dati di questo genere che ha rivoluzionato la costruzione degli strumenti di gestione dei dati e l'approccio culturale ad esso.

Per comprendere meglio di che cosa parliamo quando utilizziamo il termine *Big data* è utile, anche in questo caso, provare ad esemplificare, elencando alcuni dei tipi di *Big data* più diffusi.

---

che lo rende facilmente recuperabile dalla tabella di origine in caso di query. La terza è il partizionamento della tabella, cioè identificare per ogni tabella un criterio di suddivisione in più parti in maniera tale da facilitare la ricerca del dato quando si interroga il database.

I database relazionali agiscono, quindi, in prima istanza normalizzando il dato per archivarlo nelle tabelle e renderlo conseguentemente disponibile a metterlo in relazione con i dati contenuti in altre tabelle. Con i *Big data*, a causa della loro particolare natura, non è possibile procedere alla normalizzazione tradizionale del dato perché non possiedono un formato che possa essere contenuto in una tabella perché esso non può essere normalizzato con metodi tradizionali. Per un maggiore approfondimento tecnico si veda A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit..

30 Ibidem.

31 Riportiamo qui un esempio interessante presente in A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit.. L'autore, inoltre, spiega anche come sia la quantità che la velocità di produzione dei *Big data* possa mettere in difficoltà i database relazionali tradizionali perché non adatti a contenere un numero così elevato di dati e a riceverne in tempo reale.

Per esempio sono *Big data* i dati provenienti dai sensori e dai rilevatori scientifici e tecnici, i dati derivanti dalle quotazioni e dalle strumentazioni finanziarie, i dati provenienti dalle rilevazioni diagnostiche e mediche, i dati generati dagli strumenti di riconoscimento ed identificazione automatica che utilizzano dispositivi elettronici e che rispondono alle interrogazioni di lettori fissi o portatili a radio frequenza ed identificano in modo univoco solo un oggetto, una persona o un animale (RFID - Radio Frequency Identification).

Questa tipologia di *Big data*, sono dati prodotti in tempo reale e in quantità molto rilevanti, ma che, nella maggior parte dei casi, costituiscono un'eccezione al carattere della eterogeneità. I dati di questo genere possono essere infatti dei dati strutturati, cioè possono essere ordinati in modo tabellare in un database tradizionale (anche se comunque potenziato per ospitare una quantità straordinaria di dati), e costituiscono, come abbiamo già accennato in precedenza, una tipologia di *Big data* che possiede due delle tre caratteristiche prima analizzate, cioè il Volume e la Velocità<sup>32</sup>.

Un altro esempio di *Big data* sono i dati provenienti dal web: pagine HTML in qualsiasi lingua, tweet, contenuti dei social network (Facebook, LinkedIn etc.); commenti sui forum e blog post; documenti in vari formati (XML, PDF, RTF, Word, Excel, PowerPoint etc.). I dati provenienti dal web sono i *Big data* più noti, i più utilizzati e sono dati che contengono tutti i caratteri prima descritti: sono dati prodotti in tempo reale e in enormi quantità, eterogenei e complessi nella composizione e nelle fonti di provenienza.

Un ultimo esempio di *Big data* sono i dati GIS (Geographic Information System), i dati geografici e di localizzazione (ad esempio i dati forniti da Google Maps, per intenderci). Questi dati, come i dati provenienti dal web, sono una tipologia di dati molto eterogenei sia nelle forme che nelle fonti di provenienza, oltre ad essere dati “voluminosi” e “veloci”<sup>33</sup>.

Questa caratterizzazione dei *Big data* come insieme di dati di grande quantità, eterogeneo e disponibile in tempo reale ci mette di fronte, quindi, ad una dimensione quantitativa del fenomeno che ci è utile per comprendere principalmente come la quantità e i caratteri dei dati siano molto differenti dal passato, e questa quantità e questi determinati caratteri mettano in seria difficoltà non solo la tecnologia tradizionale di

---

32 Interessante è l'esempio riportato da A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit., di una tipologia di database chiamato database *historian* che avrebbe le capacità tecniche per archiviare una notevole quantità di dati, anche prodotti in tempo reale. Questi database sono capaci di comprimere notevolmente i dati in arrivo, anche se continuano ad avere enormi difficoltà ad archiviare dati non strutturati e richiedono un notevole investimento hardware.

33 Sul punto riprendiamo gran parte della esemplificazione della tipologia di *Big data* compiuta da A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit..

gestione dei dati, in quanto la capacità di archiviazione, anche potenziata, dei dati, e la capacità dei database relazionali di elaborarli non riescono più a fare fronte alla quantità, alla velocità di produzione dei dati e alla loro eterogeneità<sup>34</sup>, ma anche tutte le dinamiche culturali e scientifiche che stanno dietro alla gestione tradizionale dei dati.

### 3. *La prima dimensione qualitativa del fenomeno: l'evoluzione tecnologica.*

La dimensione quantitativa del fenomeno dei *Big data* che abbiamo descritto nel paragrafo precedente, e che ha cominciato a modificare la quantità e i caratteri dei dati che possiamo avere a disposizione e che sta rendendo in gran parte obsolete le tecnologie che venivano utilizzate per la loro archiviazione e per la loro elaborazione, non è l'unica dimensione rilevante del fenomeno.

L'aspetto più innovativo del fenomeno dei *Big data* è legato alla sua dimensione qualitativa. La comparsa di nuovi e numerosi dati e l'obsolescenza delle tradizionali tecnologie di archiviazione e di elaborazione dei dati sono stati tra i fattori che hanno portato ad una evoluzione innanzitutto tecnologica, che è riuscita a rimediare alle incapacità tecniche delle tecnologie tradizionali. Il primo aspetto rilevante della dimensione qualitativa del fenomeno dei *Big data*, quindi, è certamente l'aspetto tecnologico.

Oltre alla necessità di adeguare le proprie tecnologie di archiviazione e di elaborazione dei dati alle nuove esigenze tecniche che i *Big data* richiedevano, fondamentale per riuscire a gestire i dati con quelle particolari caratteristiche, negli anni 2000 si affiancò anche la necessità di conoscere i problemi e le criticità della propria azienda, di valutarne la produttività, la tenuta sul mercato e di confrontarla con la produttività e la tenuta sul mercato delle aziende concorrenti. Si cominciò, in maniera sempre più diffusa, quindi, ad analizzare ed elaborare i dati in proprio possesso per ricavare informazioni sui processi e sulle dinamiche aziendali, che descrivessero nel dettaglio lo stato di tutti i fattori rilevanti dell'azienda, anche ricorrendo alle analisi storiche dei dati. Negli stessi anni, però, almeno secondo coloro che si sono occupati della materia, parallelamente alla volontà di compiere delle analisi descrittive attraverso l'utilizzo dei dati, si è affiancata la volontà di utilizzare i dati per produrre delle analisi previsionali,

---

34 L'elaborazione dei dati che è possibile fare con i database relazionali tradizionali è soltanto un'analisi storica dell'insieme dei dati a disposizione o al massimo una analisi su ciò che sta accadendo adesso. Questo è quello che permettono di fare i *data warehouse*, che sono dei database che contengono dati integrati, provenienti da tutti i settori di una determinata organizzazione (un'azienda, per esempio). Per approfondire A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit..

con lo scopo di assicurare, a coloro che le utilizzavano, la possibilità di anticipare gli eventi e, quindi, di migliorare il processo decisionale e, soprattutto, avere un vantaggio sulla concorrenza<sup>35</sup>. Questo è avvenuto attraverso le tecnologie di *data mining*, cioè attraverso tecnologie che permisero, per la prima volta, di estrarre da un insieme di dati delle informazioni, dei risultati sintetici, delle correlazioni, che non erano immediatamente visibili e non erano predeterminabili a priori.

Una tecnologia, quella del *data mining*, che cominciava a rendere disponibili non solo delle analisi descrittive, ma anche delle prime analisi predittive e che, nel tempo, ha favorito la crescita del fenomeno della *Business Intelligence*<sup>36</sup>. La *Business intelligence*, avendo come obiettivo principale quello di raccogliere informazioni per migliorare il processo decisionale aziendale, cominciò a fare (e tutt'ora fa) due operazioni: archiviare i dati - cioè ordinare i dati aziendali raccolti (e stiamo parlando già di una rilevante quantità di dati, poiché si tratta di tutti i dati presenti in azienda) e scartare quelli che non rispettano i requisiti iniziali di programmazione, attraverso i c.d. processi di *data quality*<sup>37</sup> - e produrre delle prime analisi previsionali dei dati (ricorrendo alle tecnologie di *data mining*). La *Business intelligence*, quindi, si trasforma in un processo di analisi descrittiva e predittiva, predeterminato negli obiettivi e svolto partendo da un insieme di dati preselezionati, attraverso un rigido processo di *data quality*. Un processo, inoltre, fortemente automatizzato, in quanto prevede l'intervento della persona, del tecnico, soltanto in fase di predeterminazione degli obiettivi e di preselezione dei criteri di *data quality*, con la intrinseca possibilità che l'utilizzatore possa non conoscere la tecnologia che sta utilizzando.

---

35 Interessante è la ricostruzione storica dell'evoluzione delle tecnologie di archiviazione e di elaborazione dei dati presente in A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit..

36 La *Business Intelligence*, nota anche come *Decision Support System* (DSS), è un sistema di modelli, metodi, processi, persone e strumenti che rendono possibile la raccolta regolare e organizzata del patrimonio dati generato da un'azienda e attraverso elaborazioni, analisi ed aggregazioni, ne permette la trasformazione in informazioni, la conservazione, la reperibilità e la presentazione in forma semplice, flessibile ed efficace tale da costituire un supporto alle decisioni strategiche, tattiche ed operative. I sistemi di *Business Intelligence* fino a quel momento soltanto capaci di effettuare analisi storiche ed elaborazioni sui dati archiviati hanno avuto maggiore diffusione grazie alle tecnologie di *data mining*, in quanto permettevano loro di cominciare a fare delle analisi previsionali sui dati a disposizione. Per una maggiore precisione si v. A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit., p. 4-5.

37 Per processi di *data quality* intendiamo quei processi sui dati che indicano quali sono i dati contenuti nei database che hanno i requisiti di qualità e quali sono, invece, quelli da migliorare. La logica sottesa a questi processi di verifica della qualità dei dati a disposizione è che questi dati, essendo analizzati ed elaborati per supportare i processi decisionali, debbano essere di una qualità accettabile al fine di non commettere errori di valutazioni quando vengono utilizzati a fini decisionali. Sugli indicatori di qualità informatici ci soffermeremo nel corso dello scritto. Per una maggiore puntualità sul processo di *data quality* si consulti A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit., p. 34.

La *Business Intelligence*, però, nonostante l'innovazione tecnologica del *data mining*, presenta degli elementi di rigidità: con la predeterminazione degli obiettivi e con la preselezione dei criteri di *data quality* c'è un forte intervento dell'operatore nella fase di preparazione dell'analisi (predeterminazione degli obiettivi), guidandone conseguentemente anche i risultati, e sui dati che verranno elaborati (preselezione dei criteri di *data quality*). Rigidità che nel tempo hanno sofferto l'avvento delle logiche portate dal fenomeno dei *Big data* sull'analisi e sulla gestione dei dati e che sono state definitivamente superate con le innovazioni tecnologiche prodotte con l'avvento dei *Big data*.

Dal 2010, infatti, l'evoluzione tecnologica non solo ha migliorato la qualità delle analisi predittive e le capacità di archiviazione dei dati, ma ha prodotto anche dei rilevanti cambiamenti al processo di analisi che si compie sui dati. Si parla in maniera sempre più crescente del fenomeno della *Business Analytics*: un sistema di tecnologie, metodi e applicazioni che utilizza modelli matematici e statistici per l'analisi dei dati e per il *data mining*, capaci di analizzare i *Big data*<sup>38</sup>. La *Business Analytics* compie le medesime operazioni della *Business Intelligence*, cioè archiviazione, analisi ed elaborazione dei dati, ma in maniera parzialmente differente. È utile, allora, soffermarci su queste differenze poiché queste costituiscono l'essenza dell'innovazione tecnologica apportata dai *Big data*.

La prima differenza riguarda le tecnologie di archiviazione dei *Big data*. La differenza fondamentale con le tecnologie precedenti è che queste nuove tecnologie di archiviazione dei dati sono, dal punto di vista strutturale, più “leggere”. Innanzitutto perché vengono utilizzate delle tecniche di indicizzazione meno strutturate, molti simili ai *tags*, e la stessa costruzione delle banche dati non ha più come obiettivo quello di creare una struttura tabellare, ma ci si affida a nuovi paradigmi non strutturati<sup>39</sup>. In secondo luogo perché si alimenta una base di dati molto più grezza, in quanto è soggetta a processi di *data quality* molto meno rigidi<sup>40</sup>: si lascia spazio alla confusione, alla inesattezza del dato, prediligendo la quantità dei dati alla loro qualità. Questo, però, è un aspetto su cui torneremo tra poco quando parleremo di cambiamenti culturali prodotti dai *Big data*, in quanto non è strettamente connesso alla tecnologia, ma piuttosto ai modi in cui questa viene utilizzata.

---

38 Si v. A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit., p. 6.

39 La costruzione delle banche dati utili per l'archiviazione e la gestione dei *Big data* non prevede una struttura tabellare. Le nuove tecnologie di archiviazione sono, appunto, noSQL, cioè senza alcuna forma tabellare. Un'altra tecnologia utile per la gestione meno rigida dei *Big data* è il *Cloud computing*, un *software* di archiviazione che viene fornito come un servizio on-line.

40 A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit., p. 34.

La seconda differenza fondamentale attiene al fatto che le tecnologie di elaborazione dei *Big data* hanno bisogno di una minore predeterminazione degli obiettivi dell'analisi dei dati. Nella *Business Analytics*, abbiamo detto, si utilizzano tecnologie e metodi di analisi che utilizzano modelli matematici e statistici per produrre analisi dei dati sempre più raffinate e con una tendenza predittiva sempre maggiore e puntuale. I modelli matematici e statistici sono capaci di elaborare i dati prelevati dal database senza alcuna predeterminazione stringente degli obiettivi, producendo risultati probabilistici inediti e, alcune volte, neanche lontanamente immaginabili<sup>41</sup>. La loro potenzialità, infatti, sta proprio nella predilezione, rispetto alle tecnologie della precedente generazione, per la correlazione dei dati a loro disposizione. In realtà la correlazione non è nuova nelle elaborazioni tradizionali, ma il binomio correlazione-*Big data* rendono le stesse metodologie di correlazioni differenti rispetto al passato<sup>42</sup>.

Correlare significa quantificare la relazione statistica tra i valori di due dati: una correlazione sarà elevata quando al variare di uno dei valori, tende a variare anche l'altro; la correlazione sarà bassa, invece, se al variare di uno dei valori, l'altro resta immutato. Queste tecnologie utilizzano nelle loro analisi e nelle loro elaborazioni delle correlazioni: un metodo che ti dice quali sono i legami che si vengono a formare in via probabilistica e tralascia le motivazioni, le cause che questi legami hanno prodotto. Esemplicando, se il fenomeno A avviene spesso in concomitanza con un altro fenomeno, il fenomeno B, per prevedere come si evolverà il fenomeno A bisogna osservare le evoluzioni del fenomeno B. Prendere infatti il fenomeno B come indicatore ottimale del fenomeno A ci aiuta a prevedere in via probabilistica come evolverà il fenomeno A<sup>43</sup>.

Un metodo di analisi che utilizza la correlazione, quindi, non prevede la predeterminazione degli obiettivi da raggiungere e non prevede neanche il raggiungimento di particolari risultati attesi.

---

41 Per gli aspetti tecnici si v. A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit., p. 7, per quelli culturali si v. V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 76-77.

42 Le metodologie che prevedono l'utilizzo delle correlazioni necessitano di processi di individuazioni di variabili, i quali sono fortemente condizionati dalle ipotesi di partenza. Le correlazioni che vengono comunemente utilizzate, infatti sono correlazioni lineari, cioè correlazioni che legano due variabili sinteticamente con una linea retta. Le correlazioni utilizzate con le nuove tecnologie che utilizzano i *Big data* sono correlazioni non-lineari, in quanto più sofisticate e con più informazioni (in cui il collegamento tra le variabili è compiuto con una curva e non con una retta). Di questo si tratterà anche nel corso del lavoro, quando si parlerà dell'evoluzione culturale che i *Big data* hanno prodotto sull'elaborazione dei dati. V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 88.

43 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., 76-77.

La terza e ultima differenza è che le tecnologie che gestiscono i *Big data* devono essere governate in fase di elaborazione ed analisi dei dati, a differenza delle tecnologie della generazione precedente. Se, da un lato, il tecnico, l'operatore non predetermina in maniera rigida gli obiettivi dell'analisi e i criteri per i processi di *data quality*, con le tecnologie che gestiscono i *Big data* l'operatore ha un ruolo centrale nella costruzione del processo di analisi, in particolare nell'impostazione degli algoritmi e dei modelli matematici: il tecnico non influenza più il risultato attraverso la predeterminazione degli obiettivi delle analisi o non gestisce la base dei dati attraverso la preselezione dei criteri dei processi di *data quality*, ma gestisce direttamente il processo di analisi, attraverso la costruzione degli algoritmi e dei modelli matematici che verranno utilizzati e attraverso la selezione dei valori dei dati da correlare, ma senza darsi alcun obiettivo da raggiungere. Il tecnico, semplicemente, osserva quali correlazioni vengono fuori e fa sì che questo avvenga senza errori procedurali<sup>44</sup>.

Le differenze tra le tradizionali tecnologie e le tecnologie che gestiscono i *Big data* rappresentano il primo vero salto qualitativo compiuto grazie al fenomeno dei *Big data*: dalla rigidità strutturale si è passati ad una maggiore flessibilità; da analisi predeterminate e tendenzialmente descrittive, che servivano ad osservare il passato o il presente, si è passati ad analisi più libere e tendenzialmente predittive, che lanciano una visione probabilistica, ma attendibile sul futuro; da processi di elaborazione e analisi automatizzate si è passati a processi di elaborazione e di analisi maggiormente governati dall'uomo.

È utile sottolineare, però, che dietro (o grazie) alla modifica tecnologica si nascondono dei cambi di paradigma culturali molto importanti, che caratterizzano il secondo aspetto della dimensione qualitativa dei *Big data*.

#### *4. L'evoluzione culturale come seconda dimensione qualitativa del fenomeno: dalla logica degli Small data alla logica dei Big data.*

Se il primo aspetto della dimensione qualitativa del fenomeno dei *Big data* è l'evoluzione tecnologica, il secondo aspetto - forse quello più rilevante e che interroga trasversalmente tutte le discipline scientifiche, anche se per aspetti differenti - è l'evoluzione culturale che è scaturita principalmente dalla dimensione quantitativa del fenomeno e, in parte dall'evoluzione tecnologica, anche se qui la questione può essere un po' più complessa, ma tendenzialmente poco rilevante ai nostri fini<sup>45</sup>.

44 A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, cit., p. 8.

45 È necessario sottolineare come se, da un lato, la preminenza dell'influenza del fenomeno dell'esplosione dei dati sull'evoluzione culturale che i *Big data* stanno producendo è chiara a coloro

Quando parliamo di evoluzione culturale, parliamo essenzialmente del nuovo approccio culturale ai dati, alla loro raccolta e alla loro elaborazione, che il fenomeno dei *Big data* ha prodotto: stiamo parlando del passaggio da una logica degli *Small data* ad una logica dei *Big data*.

Secondo coloro che a oggi si sono occupati più di tutti del fenomeno dei *Big data*<sup>46</sup>, il cambiamento culturale fondamentale è riconducibile alla presa d'atto che stiamo passando da una logica di governo dei dati basata sugli *small data*, cioè basata su una quantità limitata di dati con caratteristiche determinate, ad una logica di governo dei dati basata sui *Big data*, cioè basata su una quantità di dati potenzialmente indefinibile e con caratteristiche molto differenti rispetto al passato.

La crescita della mole dei dati a disposizione e disponibili, la capacità di trarre da questi dati una grande quantità di informazioni sta modificando uno dei presupposti su cui si basava lo studio e il governo dei dati: studiare, archiviare e analizzare un insieme relativamente piccolo di dati, in quanto esemplificativo della più complessa e variegata realtà delle cose. Da una logica per cui era necessario valorizzare i dati che si erano raccolti, in quanto rappresentativi della realtà, fondamentalmente perché era possibile raccogliere soltanto una quantità limitata di dati rispetto alla realtà delle cose, si è passati ad una logica per cui è necessario raccogliere tutti i dati possibili, oramai grossomodo corrispondenti alla realtà delle cose, fondamentalmente perché è possibile farlo in maniera sempre più costante.

Il passaggio dalla logica degli *Small data* alla logica dei *Big data*, che rappresenta un cambio paradigmatico relevantissimo, è osservabile in tutti gli aspetti di gestione dei dati: dalla raccolta, alle finalità del loro utilizzo, passando per le modalità della loro elaborazione.

Per molti versi i cambiamenti culturali che stanno coinvolgendo i vari aspetti della gestione dei dati non sono del tutto nuovi, non sono fenomeni nati con i *Big data*, sono fenomeni conoscitivi già presenti nella logica degli *Small data*, ma che si sono profondamente rinnovati con il fenomeno dei *Big data*: si sono rinnovati per l'estensione che hanno raggiunto, per le modalità con cui vengono prodotti e per la crescente intensità del loro impatto sulla società.

---

che si sono occupati della materia, come V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., più controverso ci pare il legame tra evoluzione tecnologica ed evoluzione culturale. La questione è quanto le nuove disponibilità tecnologiche, che hanno concretizzato la possibilità di gestire i *Big data*, abbiano influenzato l'evoluzione culturale o quanto l'evoluzione culturale abbia incentivato la ricerca di nuove soluzioni tecnologiche per elaborare i *Big data*. Una questione di difficile risoluzione, che ci sembra essere risolvibile con una interpretazione che integri entrambe le ipotesi.

46 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit..

È utile allora sottolineare in modo più puntuale come la logica dei *Big data* abbia modificato l'archiviazione, l'elaborazione e le finalità di utilizzo dei dati.

#### 4.1. La datizzazione della realtà.

La nuova logica dei *Big data* ha coinvolto innanzitutto l'approccio classico alla raccolta dei dati. Si è passati, ovviamente con dei processi gradualmente, da una logica di raccolta del dato circoscritta, alla vera e propria datizzazione della realtà. La datizzazione della realtà, non è certo un fenomeno nuovo: l'uomo ha sempre avuto la tendenza a raccogliere i dati sulla realtà che lo circondava, si è sempre dotato di mezzi di misurazione e di strumenti di archiviazione per conoscere la realtà in cui viveva, mezzi che, tendenzialmente, gli garantivano una circoscritta quantità di dati rispetto alla realtà del caso<sup>47</sup>. Per datizzazione della realtà si intende, però, quella tendenza a convertire qualsiasi fenomeno, qualsiasi cosa che sia reale, in forma quantitativa, in dato, in modo da poterlo archiviare e, soprattutto, analizzare<sup>48</sup>. Con il fenomeno dei *Big data*, è migliorata notevolmente la capacità tecnologica di misurare e di archiviare la realtà, come abbiamo avuto modo di vedere, ma principalmente è cresciuta enormemente la volontà di misurare e archiviare dati sulla realtà, anche sui fenomeni più impensabili, come gli stessi comportamenti e le stesse relazioni tra persone.

Tra i fenomeni più recenti è utile, per esemplificare il fenomeno della datizzazione, indicare quello di Google Books e della sua capacità di indicizzare tutte le parole di tutti i testi digitalizzati e contenuti al suo interno, tanto da poter produrre dei grafici sull'utilizzo di determinate parole o di determinate frasi nel tempo. Questa capacità di datizzare le parole e le frasi ha fatto nascere anche una disciplina, la culturomica, una lessicologia informatizzata che tenta di comprendere il comportamento umano e i trend culturali tramite le analisi quantitative del testo<sup>49</sup>. Un altro esempio di datizzazione è la geolocalizzazione dei luoghi, delle persone e degli oggetti. Il noto GPS (Global Positioning System), strumento che già permetteva da tempo di geolocalizzare i luoghi e gli oggetti, è, oggi, solo uno dei modi con cui farlo: oramai il dialogo tra GPS, reti wireless, reti di telefonia e apparecchi *mobile*, non solo rende la geolocalizzazione dei

---

47 Per un'approfondita ricostruzione storica sulla tendenza dell'uomo a datizzare la realtà si v. A. W. CROSBY, *The Measure of reality*, Cambridge University Press, Cambridge, 1996, tr. it., *La misura della realtà*, Dedalo, Bari, 1998.

48 Questa definizione di datizzazione è presente in V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 103-130.

49 Per i due esempi appena riportati si v. V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 116-117.

luoghi e degli oggetti, dettagliata e precisa, ma rende possibile anche la geolocalizzazione delle persone (si pensi alle applicazioni e gli strumenti di localizzazione degli smartphone). Gli strumenti di geolocalizzazione sui veicoli e sulle persone sono stati utilizzati, per esempio, dalle compagnie assicurative per modificare la valutazione del rischio in quanto ora è possibile analizzare nel dettaglio i tempi di percorrenza, i luoghi attraversati e i comportamenti individuali. UPS, invece, riesce ad utilizzare i geolocalizzatori dei suoi veicoli in vari modi, tra cui razionalizzare i percorsi di trasporto, controllare lo stato dei veicoli e il comportamento dei dipendenti<sup>50</sup>.

In ultimo è utile sottolineare come ci sono esempi anche di datizzazione del comportamento dell'uomo o delle sue relazioni sociali. Del primo tipo è il c.d. *reality mining*, che consiste nel processare enormi quantità di dati provenienti dai cellulari per trarne delle inferenze e previsioni sul comportamento umano, mentre del secondo tipo sono i grafici sociali, cioè gli schemi relazionali che Facebook, per esempio, costruisce costantemente con i nostri contatti e con tutte le operazioni che facciamo sulla piattaforma<sup>51</sup>.

Dagli esempi appena compiuti è possibile evincere come oramai si ha la tendenza culturale a datizzare ogni cosa e chiunque. La datizzazione della realtà rispecchia perfettamente il cambiamento culturale che si sta diffondendo in merito alla logica che governa l'archiviazione dei dati ed è un paradigma con cui chiunque conserva e gestisce i dati deve confrontarsi, sia nel settore privato, e gli esempi riportati dimostrano come questo stia già avvenendo, sia nel settore pubblico.

#### 4.2. Dal “perché” al “che cosa”: osservare (e non spiegare) e prevedere i fenomeni.

Il passaggio dalla logica degli *Small data* alla logica dei *Big data* sta coinvolgendo anche le modalità di elaborazione e analisi dei dati. Con le considerazioni che si sono fatte sull'evoluzione tecnologica si sono già anticipati due aspetti che ci sono molto utili nella delineazione dell'evoluzione culturale intercorsa sulle modalità di elaborare e di analizzare i dati: il primo è legato ai processi di *data quality*, quando abbiamo sottolineato come quei processi, quando parliamo di *Big data*, siano nettamente più

---

50 Per Entrambi gli esempi sulla geolocalizzazione si v. V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 123-124.

51 Sul *reality mining* e sui grafici sociali si v. V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 127-130.

leggeri o addirittura assenti rispetto a tempi meno recenti; il secondo è legato ai metodi statistici e matematici che caratterizzano le tecnologie dei *Big data*, i quali utilizzano la correlazione dei valori dei dati come principale tecnica di elaborazione.

Questi due elementi mettono in luce in maniera molto più chiara un approccio culturale all'elaborazione e all'analisi dei dati molto differente rispetto all'era degli *Small data*.

Con la possibilità di reperire soltanto una certa quantità dei dati rispetto alla complessità dei fenomeni che si studiavano e con la conseguente necessità che questi dati fossero esatti, l'approccio culturale all'elaborazione e alle analisi dei dati metteva al centro di questi processi i metodi di campionamento, in risposta alla prima necessità, e la preventiva selezione dei dati da archiviare per garantirne la qualità, in risposta alla seconda necessità. Con i *Big data*, sia il campionamento sia l'attenzione alla qualità dei dati, diventano recessivi. Il campionamento, evidenza metodologica della volontà/possibilità di studiare una parte per comprendere il tutto, appartenente alla logica degli *Small data*, comincia ad essere sostituito dalla possibilità/volontà di studiare il tutto per comprendere il tutto, cioè dalla possibilità/volontà di raccogliere tutti i dati che è possibile raccogliere su un determinato fenomeno al fine di studiarlo, dati che presumibilmente si avvicinano molto alla totalità dei dati ricavabili. L'attenzione per la qualità dei dati, allo stesso modo, legata necessariamente alla volontà/possibilità di utilizzare pochi dati, ma i più esatti possibile, viene sostituita dalla possibilità/volontà di non selezionare più i dati con processi che valutino la qualità del dato al fine di non intaccare la massa quantitativa necessaria per le analisi o per le elaborazioni, le quali soltanto utilizzando tutti i dati possibili possono produrre dei risultati probabilistici di alta qualità, risultati che non possono essere conseguiti con i metodi tradizionali<sup>52</sup>.

Un esempio può chiarire meglio la differenza di approccio all'esattezza e alla qualità dei dati nella logica dei *Big data*. Se all'interno di un vigneto volessimo misurare la temperatura interna per mantenerla costante, la logica degli *Small data* ci porterebbe a installare un solo sensore e avremmo la necessità che esso trasmetta dati esatti e che non ci siano errori di taratura. Con la logica dei *Big data*, invece, potremmo predisporre dei sensori per tutti i filari della vigna in maniera tale da avere molti più sensori, economicamente meno onerosi magari, che trasmettono dati con una percentuale di errore maggiore e con un tasso di incertezza certamente più alto, compensato, però, da un numero di dati molto maggiore e una visione più capillare della situazione. La logica dei *Big data* ci porterebbe a fare un ulteriore passo in avanti. Se passassimo da una rilevazione al minuto ad una decina di rilevazioni al secondo, la situazione diventerebbe, da un lato, notevolmente confusa perché l'accuratezza della sequenza di

---

52 In questo senso in maniera convincente V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit.

registrazione verrebbe meno (magari delle rilevazioni possono arrivare in ritardo e quindi sfasarla), ma, dall'altro, notevolmente fruttuosa in quanto la mancanza di accuratezza dei dati verrebbe ampiamente compensata dalla possibilità di moltiplicare all'infinito la mole di informazioni che è possibile reperire e dalla possibilità di elaborare questi dati per prevenire qualsiasi tipo di disagio<sup>53</sup>.

Più informazioni a disposizione, significa migliori risultati. A dimostrazione di ciò è necessario riportare due ulteriori esempi su come il semplice aumento della quantità dei dati possa migliorare i risultati legati al loro utilizzo. Il primo esempio emblematico è il miglioramento repentino del sistema di traduzione di Google. Secondo gli esperti, semplicemente aumentando di decine di migliaia di volte il proprio *dataset* di riferimento, rispetto ad altri sistemi di traduzione (come Candide, il sistema di traduzione dell'IBM), attraverso i dati provenienti dalle pagine web e dalla digitalizzazione dei testi, Google ha potenziato enormemente il suo sistema di traduzione rendendolo capace di tradurre 60 lingue diverse e di mettere in relazione lingue diversissime come l'hindi e il catalano, che forse nessuno ha mai tradotto davvero<sup>54</sup>. Questo è stato possibile perché Google trattava il linguaggio come una massa caotica di dati molto grezzi. Il *dataset*, infatti, conteneva errori ortografici, frasi senza un senso logico, a cui Google ha applicato delle inferenze probabilistiche hanno drasticamente diminuito i margini di errore proprio perché era un *dataset* contenente un'enorme quantità di dati. La confusione dei dati è stata in questo caso fondamentale per migliorare i risultati che ci si era prefissati.

Un altro esempio emblematico è l'evoluzione della classificazione dei contenuti. La perfetta ed esatta collocazione di tutto quello che un archivio contiene secondo le regole della tassonomia, tipico approccio della logica degli *Small data*, entra in enorme difficoltà con l'aumento repentino ed esponenziale dei dati, tant'è che si stanno sviluppando tecniche di classificazione molto più flessibili, più disordinate, come il *tagging*, che oramai si è imposto come il metodo di classificazione più diffuso su Internet<sup>55</sup>.

Nell'epoca dei *Big data*, quindi, la confusione è un elemento ineludibile della realtà. Qualcuno ha sottolineato anche come questo approccio ai dati rispecchi e sia più rispondente alla realtà che ci circonda, una realtà tendenzialmente caotica e articolata,

---

53 L'esempio del vigneto come emblema del passaggio dalla logica degli *Small data* alla logica dei *Big data* è riportato in V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit..

54 A. HAVELY - P. NORVING – F. PEREIRA, *The Unreasonable effectiveness of data*, in *IEE Intelligence System*, marzo-aprile 2009, pp. 8-12.

55 Taggare significa classificare dei contenuti mettendoci un'etichetta. Gli utenti utilizzano dei tags, delle etichette appunto, senza un particolare criterio tassonomico da rispettare per procedere con il tag. Si v. V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 65.

mai corrispondente alla logica ordinatrice che ha ispirato l'approccio ai dati fino ad ora<sup>56</sup>. L'esattezza delle rilevazioni e la qualità dei dati entrano in crisi e lasciano spazio all'inesattezza, alla confusione, al disordine per un motivo fondamentale: aumentare la quantità di dati a disposizione, senza alcun tipo di filtro, e migliorare la credibilità delle proprie previsioni.

Nella logica dei *Big data*, la recessività dell'esattezza e dell'attenzione alla qualità dei dati porta anche ad una preferenza di alcuni metodi di analisi dei dati rispetto ad altri e ad un mutamento della *ratio* e degli obiettivi dello studio e dell'elaborazione dei dati.

Si passa da un'analisi accurata, alla ricostruzione dei trend generali dei fenomeni. Nella logica degli *Small data*, caratterizzata da un penuria di dati o comunque una quantità limitata di dati, per analizzare un fenomeno si procedeva a costruire un'ipotesi, a scegliere le conseguenti variabili con i relativi indicatori, avendo come obiettivo principale l'individuazione delle cause dei fenomeni. I metodi di analisi erano differenti e le correlazioni che vengono utilizzate, a causa di scarse capacità di elaborazione, erano correlazioni lineari, coerenti con la logica per cui l'obiettivo finale è semplificare la realtà e cercare i rapporti di causa-effetto al fine di spiegare i fenomeni. Con la logica dei *Big data*, il paradigma e gli strumenti mutano. La disponibilità di una grande quantità di dati e la possibilità di elaborarli attraverso l'utilizzo di correlazioni non lineari e non-causali, rende inutile fare delle ipotesi e scegliersi le variabili, in quanto è possibile avere degli indicatori ottimali che sono già il frutto della prima elaborazione dei dati, con l'obiettivo non di ricercare rapporti di causa-effetto, di capire il perché dei fenomeni, ma di individuare che cosa sta accadendo, delle “piste” da seguire, di evidenziare dei legami tra variabili che prima, basandoci soltanto su delle ipotesi precostituite, non avremmo neanche pensato. Le correlazioni ci portano a comprendere dinamiche tecniche e sociali che non avremmo mai compreso.

In effetti, l'elemento dell'imprevedibilità e dell'originalità delle relazioni create attraverso le correlazioni non lineari e non-causali su un *dataset* molto rilevante, è evidente se si riporta qualche esempio.

La catena di discount statunitense Target ha utilizzato un team di analisi per calcolare un indice di “previsione della gravidanza”, cioè un indice che fosse in grado di prevedere quali clienti aspettassero un figlio a breve, attraverso i loro modelli di acquisto, così da attivare delle campagne di marketing mirate. Le correlazioni dei dati raccolti sugli acquisti delle donne attraverso le iscrizioni ad alcuni registri settoriali (come i registri per i regali ai bambini 0-6 anni etc.) hanno permesso alla catena di

---

56 Si v. diffusamente D. WEINBERGER, *Everything is miscellaneous: the power of the new digital disorder*, Henry Holt and Company, New York, 2008, tr. it., *Elogio del disordine. Le regole del nuovo mondo digitale*, Rizzoli, Milano, 2010.

discount di stimare con ragionevole attendibilità anche la data del parto e di mandare buoni sconto direttamente a casa per l'acquisto di determinati prodotti<sup>57</sup>.

Un altro esempio delle potenzialità della correlazione è legato al settore sanitario. Un team di ricercatori dell'Institut of Technology dell'Università dell'Ontario e dell'IBM ha costituito un software in grado di raccogliere dati biometrici sui neonati prematuri per prevenire malattie infettive. Utilizzando le correlazioni è stato possibile verificare come prima di un'infezione i parametri biometrici siano tutti stabili, un aspetto che va contro la logica convenzionale medica, e che ha permesso ai medici di non considerare la stabilità dei parametri come un segnale positivo<sup>58</sup>.

Un ultimo esempio dell'utilizzo di correlazioni originali è legato alla verifica della stabilità delle temperature dei tombini newyorchesi. La società pubblica che fornisce corrente elettrica, Con Edison, è riuscita attraverso la correlazione di tutti i dati in suo possesso a rilevare le due variabili principali utili per prevedere il cedimento delle coperture di ghisa dei tombini della città, provocandone l'esplosione: l'età e i tombini soggetti a precedenti incidenti<sup>59</sup>.

Come è possibile vedere anche dagli esempi, un altro elemento fondamentale dell'elaborazione dei *Big data* è la predittività, ovviamente probabilistica, delle sue analisi. L'elaborazione con metodi statistici e matematici riesce a garantire una certa capacità di previsione attendibile di quello che potrebbe accadere, soprattutto perché agisce su una grande mole di dati. Oltre agli esempi appena fatti, in cui la grande quantità di dati raccolti sui neonati prematuri aveva permesso ai medici non solo di individuare quali variabile tenere sotto osservazione, ma anche chi dei prematuri sarebbe stato più a rischio di infezione, in maniera tale da poter intervenire tempestivamente qualora si fosse realmente verificato, così come la grande quantità di dati raccolti sullo stato della rete elettrica della città di New York aveva permesso non solo l'individuazione delle variabili che bisognava osservare per prevenire l'esplosione dei tombini della città, ma anche quali tombini fossero a rischio di esplosione, un altro esempio vale la pena di citare in questo scritto.

Nel 2009 Google ha pubblicato un saggio in cui descriveva un esperimento compiuto dai propri ingegneri in cui questi erano stati capaci di prevedere la diffusione della

57 Questo esempio è ripreso da C. DUHIGG, *The power of habit: why we do what we do in life and business*, Random House, New York, 2012, tr. it., *La dittatura delle abitudini*, Corbaccio, Milano, 2012.

58 C. MCGREGOR - C. CATLEY - A. JAMES - J. PADBURY, *Next Generation Neonatal Health Informatics with Artemics*, in European Federation for Medical Informatics, A. MOEN ET ALTRI (A CURA DI), *User Centred Networked Health Care*, IOS Press, Lansdale, 2011.

59 Si v. C. RUDIN ET AL., *21st-Century Data Miners Meet 19th Electrical Cables*, in "Computer", giugno 2011, p. 103-105. Un riferimento più tecnico è C. RUDIN ET AL., *Machine Learning for the New York City Power Grid*, in IEEE Transactions on Pattern Analysis and Machine Intelligence, 34, n. 2 (2012), pp. 328-345.

classica influenza invernale negli USA, sia a livello federale sia a livello dei singoli Stati, attraverso l'elaborazione dell'oggetto delle *query* (delle parole digitate sul motore di ricerca) inserite dagli utilizzatori di Google<sup>60</sup>: stiamo parlando di più di 3 miliardi di *query* ogni giorno, archiviate ed elaborate dalla multinazionale. Gli ingegneri di Google dopo una prima analisi dei dati, hanno selezionato 50 milioni di parole più utilizzate e li hanno confrontati con i dati sulla diffusione dell'influenza invernale tra il 2003 e il 2008 diffusi dalle autorità sanitarie pubbliche, cioè i Centers for Disease Control and Prevention (CDC). La semplice correlazione tra i dati relativi alle frequenze di determinate *queries* e i dati sulla diffusione dell'influenza invernale nel tempo e nello spazio, senza nessuna ipotesi iniziale, ha portato a dei risultati molto interessanti: il software ha scoperto una combinazione di 45 parole-chiave che, quando venivano utilizzate insieme nei modelli matematici presentavano un forte correlazione tra la loro previsione di diffusione dell'influenza e i dati ufficiali relativi all'intero territorio nazionale che quella influenza aveva registrato davvero. Queste correlazioni avrebbero potuto dire, qualora fossero state utilizzate, dove si era propagata l'influenza in tempo reale, a differenza dei dati dei CDC che hanno ricostruito tutta la dinamica di diffusione con almeno due settimane di ritardo<sup>61</sup>.

Con la logica dei *Big data*, quindi, si tentano di superare i limiti della causalità, che ha dominato lo studio e l'analisi dei dati fino a questo momento, e di valorizzare le potenzialità delle predizioni probabilistiche che alcune modalità di analisi dei dati permettono di fare.

Ricerca delle connessioni causali per la spiegazione dei fenomeni, ricerca le motivazioni di fondo che guidano le dinamiche che caratterizzano quei fenomeni, è una tendenza, non solo dovuta a fattori cognitivi<sup>62</sup>, ma anche, o soprattutto, culturali. Un approccio che vede i suoi limiti nella limitatezza dei dati a disposizione, che porta l'analista a fare delle ipotesi sui nessi di causalità tra delle variabili senza una verifica approfondita; nella necessità di indirizzare l'analisi dei dati, che rischia di far perdere la visione complessiva del fenomeno o alcuni aspetti che non sono strettamente collegati alle ipotesi di partenza; nella presenza di margini di errore, nonostante si stia parlando di metodi causali lenti e metodici; nel tempo che dimostrare una causalità tra delle variabili richiede.

---

60 J. GINSBURG ET AL., *Detecting Influenza Epidemics Using Search Engine Query Data*, Nature, 457, 2009, pp. 1012-1014.

61 A.F. DUGAS ET AL., *Google Flu Trend: Correlation with Emergency Department Influenza Rates and Crowding Metrics*, CID Advanced Access, 8 gennaio 2012.

62 Interessante in questo senso è la differenziazione che fa D. KANHEMAN, *Thinking Fast and Slow*, Farrar, Straus and Giroux, New York 2011, pp. 74-75, tr. it., *Pensieri lenti e veloci*, Mondadori, Milano, 2012, in merito alla causalità c.d. immaginaria e la causalità frutto del lavoro di analisi lento e metodico

Le correlazioni non causali e non lineari, infatti, presuppongono l'abbandono della logica della causalità, e permettono di superare i limiti appena elencati: agiscono su un bacino di dati enorme, producendo delle analisi probabilistiche e predittive che hanno un margine di errore accettabile; liberano l'analisi dei dati dagli indirizzi iniziali delle ipotesi e lasciano “parlare i dati”, divenendo anche un'ottima base di partenza per ipotetiche e successive ipotesi causali, a questo punto molto più mirate ed efficaci; sono strumenti molto più veloci ed immediati rispetto alle analisi causali.

Le correlazioni, inoltre, sui *Big data* hanno una spiccata capacità di “prevedere” probabilisticamente l'evoluzione dei fenomeni, di avvicinarsi molto e in maniera tendenzialmente attendibile a quello che potrebbe accadere, di dare delle chiavi di lettura dei fenomeni originali e differenti rispetto alle ipotesi che saremmo tentati ragionevolmente di fare. Una potenzialità che solo essi posseggono.

È utile, però, sottolineare come entrambi gli elementi, il superamento della causalità e le “predizioni” probabilistiche abbiano suscitato un dibattito acceso tra gli studiosi. Alcuni di questi hanno messo in evidenza i limiti di queste tipologie di previsioni, ridimensionandone la portata innovativa. Le condizioni iniziali di qualsiasi fenomeno, secondo questi studiosi, sono soggette a degli errori infinitesimali che diventeranno esponenzialmente più grandi durante lo svolgimento del fenomeno corrompendo anche in maniera rilevante la previsione stessa: esiste, quindi, un orizzonte di predicibilità, oltre il quale non è possibile fare previsioni attendibili e questo orizzonte diventa sempre più stretto in ambiti scientifici dove non agiscono delle leggi deterministiche (come l'ambito giuridico, politico, economico e sociale)<sup>63</sup>. I *Big data*, quindi, restano, per coloro che hanno questa interpretazione, una ulteriore opportunità di indagine a cui bisogna approcciarsi senza alcun pregiudizio iniziale, e quindi sfruttarne le potenzialità, ma anche senza troppe aspettative<sup>64</sup>.

Questo discorso sul superamento della causalità, invece, ha portato alcuni osservatori a sottolineare come la fine delle ipotesi e della causalità possa portare al tramonto del tradizionale metodo scientifico, che vede nella causalità e nelle ipotesi i suoi presupposti di partenza, e, più in generale, alla fine della teoria<sup>65</sup>. Questi approcci, anche provocatori per stessa ammissione degli autori<sup>66</sup>, sostenevano che il processo tradizionale su cui si fondava la scoperta scientifica - formulare un'ipotesi e cercarne la dimostrazioni con metodi che avevano come obiettivo la dimostrazioni di rapporti di causa-effetto – era destinato a scomparire a favore delle nuove analisi statistiche e

---

63 F. SYLOS LABINI, *Big data, complessità e metodo scientifico*, in *Aspenia*, n. 63, Aspen Institute Italia, dicembre 2013.

64 Ibidem.

65 C. ANDERSON, *The Petabyte Age*, in *Wired*, giugno 2008.

66 Lo stesso Chris Anderson sulla National Public Radio all'interno del programma *Search and Destroy* il 18 luglio 2008

matematiche dei dati, totalmente svincolate dalle teorie: l'analisi dei dati non avrebbe bisogno di nessuna base concettuale.

Ragionevolmente, però, si può facilmente essere d'accordo con chi in più sedi ha sostenuto come la “fine della teoria” sia una forzatura poco aderente alla realtà delle cose<sup>67</sup>. Innanzitutto perché gli strumenti di analisi dei *Big data* sono strumenti statistici, matematici e informatici che si fondano su quelle medesime teorie che si vogliono superare e soprattutto perché l'interpretazione dei risultati di quelle stesse elaborazioni potranno essere utilizzate a fini scientifici, politici, sociali, giuridici ed economici prevalentemente attuando teorie già esistenti o costituendo la base di nuove teorie scientifiche (strettamente intese), economiche, giuridiche, politiche e sociali.

#### 4.3. *L'estrema valorizzazione dei dati e le molteplici finalità del loro utilizzo.*

La logica dei *Big data*, in ultimo, ha modificato anche l'approccio alle finalità dell'utilizzo dei dati e alla loro valorizzazione. L'approccio tradizionale ai dati prevedeva che questi venissero raccolti unicamente per predeterminate finalità, stabilite fin dall'inizio e, una volta utilizzato il dato per quelle medesime finalità, esso veniva accantonato e mai più utilizzato o, nel peggiore dei casi, eliminato. Con la logica degli *Small data* il c.d. «valore opzionale» dei dati<sup>68</sup>, cioè il valore che si può ricavare da tutti i modi in cui è possibile impiegarli, non viene quasi mai utilizzato a pieno, in quanto il dato si raccoglie e si utilizza per una sola finalità. Con la logica dei *Big data*, invece, la tendenza, già presente da tempo, a riutilizzare i dati per finalità diverse da quelle primarie, da quelle cioè che hanno portato alla raccolta del dato, la tendenza a utilizzare il «valore opzionale» dei dati, è cresciuta in maniera esponenziale ed è diventata quasi culturalmente il presupposto necessario per il funzionamento della logica dei *Big data*. In tutti i casi di utilizzo dei *Big data* che abbiamo riportato fino ad ora, infatti, è stato possibile osservare come il riutilizzo dei dati per finalità differenti rispetto a quelle primarie sia stato centrale, come nel caso dell'utilizzo delle *query* di Google per individuare la diffusione delle epidemie influenzali. Ci sono stati casi in cui la ricombinazione dei dati, cioè l'utilizzo combinato di *dataset* differenti, è stato determinante per creare valore ulteriore e più grande rispetto alla semplice somma dei

---

67 D. BOYD – K. CRAWFORD, *Six Provocations for Big Data*, presentato al convegno dell'Oxford Internet Institute “A Decade in Internet Time: Symposium on the Dynamics of Internet and Society”, 21 settembre 2011; V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 99-101; F. SYLOS LABINI, *Big data, complessità e metodo scientifico*, cit..

68 Così viene definito in V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp.140-143.

valori dei dati presi singolarmente, come è successo nel caso dei tombini di New York. La logica dei *Big data*, però, relativa alla maggiore valorizzazione possibile dei dati trova la sua principale forma di espressione nella possibilità/volontà di reperire i dati già con l'intento di utilizzarli per più finalità e quindi archivarli in formati e con modalità che garantiscano loro di essere estensibili, appunto utilizzabili da subito per più di una finalità<sup>69</sup>. Passare dal reperire i dati per una sola finalità al reperire i dati avendo già in mente più modalità del loro utilizzo è certamente uno degli elementi fondanti la logica dei *Big data* relativamente alla valorizzazione dei dati, anche se è probabile che lo sviluppo tecnologico o una diversa attenzione alle possibilità di utilizzo dei dati a propria disposizione possono fare emergere nuove modalità di utilizzo dei dati difficilmente percepibili al momento della loro raccolta.

Riutilizzo, ricombinazione e fusione dei *dataset*, strumenti e metodi per rendere estensibili i dati sono, quindi, le tre tipologie di valorizzazione dei dati che hanno moltiplicato le finalità del loro utilizzo.

La logica dei *Big data* non ha soltanto evidenziato le potenzialità della moltiplicazione dell'utilizzo dei dati raccolti, ma ha anche estremizzato la valorizzazione dei dati. Abbiamo già avuto modo di dire come nella logica dei *Big data*, anche i dati errati e di cattiva qualità hanno un loro valore, non solo in quanto contribuiscono alla formazione di quel dato quantitativo determinante per potere parlare di *Big data*, ma anche perché possono essere a loro volta utilizzati, come ha fatto la Microsoft, per migliorare il correttore automatico di Office Word, o come ha fatto in maniera molto più fruttuosa Google, che ha utilizzato i dati errati per migliorare il completamento automatico della ricerca, i propri servizi Gmail e GoogleDocs e il suo sistema di traduzione automatica<sup>70</sup>. Oltre ai dati erronei o di cattiva qualità, la logica dei *Big data* prevede anche l'utilizzo dei c.d. dati residui, i dati cioè che si creano come sottoprodotto delle azioni o degli spostamenti che si compiano sulla rete<sup>71</sup>. Ogni azione che viene compiuta in rete – la posizione del link che apriamo dopo una ricerca, il tempo di permanenza su una pagina - produce potenzialmente dei dati che possono essere archiviati e analizzati per molteplici finalità. Google ha utilizzato i dati sul numero delle volte che una determinata persona ha ricercato un termine o suoi correlati, i dati che quantificavano la frequenza con cui determinate persone hanno cliccato su una pagina o i dati sulla posizione del link selezionato più frequentemente per migliorare costantemente il suo motore di ricerca secondo le preferenze degli utenti<sup>72</sup>. Facebook ha utilizzato i dati residui sulla reazione dei propri utenti ai post dei loro amici sulla piattaforma (più gli amici su Facebook

---

69 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp.149-150

70 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p.155.

71 Ibidem

postano qualcosa, più si è tentati di postare) per rimodellare la struttura della Homepage e mettere maggiormente in evidenza le attività dei propri amici. Anche fuori dal settore informatico i dati residui possono essere valorizzati: per esempio le aziende che producono e-book, come Amazon, possono utilizzare i dati su che cosa leggono i lettori, sulle loro annotazioni al testo, sul tempo che un lettore trascorre su una pagina, se la lettura è fugace o se dopo avere letto qualche pagina il testo non viene più aperto, tutto per raccogliere più feedback possibili e migliorare i propri servizi<sup>73</sup>.

Nonostante il valore dei dati stia diventando, come è possibile osservare, enorme e in crescente aumento, bisogna sottolineare, per avere una visione completa del valore dei dati al tempo dei *Big data*, come la capacità di produrre dati di queste caratteristiche comporti anche la possibilità che l'insieme dei dati raccolti possa perdere valore in maniera repentina. I dati possono perdere valore per una serie di motivi, primo tra tutti perché troppo datati. Fare analisi di qualsiasi genere con dei dati non aggiornati rischia di distorcere l'elaborazione e di produrre dei risultati poco utili, e fare analisi con dei *Big data* non aggiornati rischia di distorcerne il risultato e di scardinare totalmente tutte le potenzialità, in primo luogo predittive, che i *Big data* garantiscono. L'attualità del dato è fondamentale nella logica dei *Big data* e deve restare una costante strutturale per utilizzare e sfruttare a pieno le potenzialità dei *Big data*.

L'aggiornamento o meno del dato, però, non è sempre un elemento fondante la perdita di valore del dato: alcuni dati, anche se non sono aggiornati possono essere ancora utilizzabili per fare analisi ed elaborazioni. Si pensi ai dati che possano avere perso un determinato valore perché non possono essere più utilizzati per alcune finalità, magari per la finalità primaria, ma che potrebbero averne ancora molto se utilizzati per finalità secondarie ed ulteriori<sup>74</sup>.

La logica dei *Big data*, quindi, ha portato con se non solo una enorme crescita della dimensione quantitativa dei dati e un cambiamento tecnologico rilevante, ma sta modificando in maniera importante anche alcuni aspetti rilevanti dello studio, dell'analisi e della percezione del valore dei dati. La datizzazione della realtà, i metodi di correlazione e le analisi predittive, la tendenza a moltiplicare le finalità di utilizzo dei dati sono tutte componenti di un più generale cambiamento culturale che non potrà che crescere nel tempo. Non tutto ciò che noi abbiamo, per finalità descrittive, incasellato

---

72 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp.156.

73 Gli esempi su Facebook e Amazon sono riportati in V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp.156.

74 Sulla necessità che i dati siano sempre aggiornati e sulla netta differenziazione tra valore del dato e il suo aggiornamento si v. ampiamente V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp.149-161.

nella logica degli *Small data*, andrà perduto, ma verrà utilizzato in maniera differente. Si pensi alla ricerca dei rapporti di causalità tra le variabili per spiegare i fenomeni, che resterà determinante per comprendere la realtà e che probabilmente verrà, come abbiamo sottolineato in precedenza, utilizzata dopo una prima analisi dei dati attraverso le correlazioni. Oppure si pensi ai processi di *data quality* e, più in generale, all'attenzione per la qualità dei dati: in alcuni settori, come quello pubblico, o per determinate necessità, la qualità dei dati resterà comunque un obiettivo da raggiungere, anche se con strumenti e in momenti differenti.

##### *5. Le tipologie di utilizzo dei Big data: potenzialità e limiti.*

Il fenomeno dei *Big data*, dunque, è un fenomeno complesso con diverse declinazioni, soprattutto culturali, che merita - per gli effetti che sta producendo e che in potenza produrrà in maniera sempre più frequente - di essere analizzato approfonditamente da tutte le discipline scientifiche, in particolar modo dalle discipline giuridiche. Il fenomeno dei *Big data* sta assumendo, in modo esponenziale, una concretezza tale che non può più essere ignorato. Le implicazioni giuridiche che l'utilizzo dei *Big data* produce sono molteplici ed attengono a svariati aspetti, che avremo modo di riportare nei successivi due capitoli. È utile allora, anche per comprendere concretamente queste implicazioni, tentare innanzitutto di elencare i modi e le tipologie dell'utilizzo che si sta facendo attualmente dei *Big data*.

Il fenomeno dei *Big data* è un fenomeno che nasce principalmente nel settore privato, come abbiamo accennato in precedenza. La prima tipologie di utilizzo è, quindi, certamente quella commerciale, sia nel senso di miglioramento delle capacità di vendita delle aziende private o, in generale, di miglioramento dei servizi esistenti o di creazione di nuovi servizi a fini commerciali, sia declinata come capacità di elaborare i *Big data* e vendere le informazioni che si sono prodotte.

Alcuni esempi del primo tipo sono già stati fatti nel corso dello scritto, come quello della catena di discount statunitense Target, la quale è riuscita a targetizzare la clientela in attesa di un bambino per avviare delle campagne di vendita mirate. Altri esempi sono molto noti e di più immediata comprensione, come la capacità di LinkedIn di indicarci quali potrebbero essere i nostri potenziali contatti o quella di Amazon di indicarci in maniera puntuale quali libri potrebbero interessarci o la stessa Google, che, utilizzando i *Big data*, riesce a comporre la lista dei risultati della ricerca basandosi sulle ricerche effettuate in precedenza e, quindi, sulle nostre preferenze.

Ci sono, però, altri esempi, che meritano di essere esplicitati. Interessante è l'esperienza di *Farecast* di Oren Etzioni che - attraverso l'elaborazione di un campione di 12.000 osservazioni sull'andamento dei prezzi per 41 giorni consecutivi e l'acquisizione del database più fornito del settore, contenente tutte le informazioni sui posti disponibili delle tratte commerciali dell'aviazione americana in un anno - è riuscito a costruire un modello previsionale sull'andamento dei prezzi dei biglietti delle compagnie aeree statunitensi, attraverso il quale fornisce dei consigli su quando procedere all'acquisto del biglietto e anche sull'attendibilità delle proprie previsioni<sup>75</sup>. *Farecast* è stata una delle prime start-up che ha utilizzato *Big data*, negli USA. Nata intorno al 2000, l'azienda di Etzioni è annoverabile certamente non solo tra le prime esperienze che hanno utilizzato i *Big data*, ma anche tra le esperienze che aveva cominciato ad utilizzare i dati con il paradigma culturale che caratterizzerà il fenomeno dei *Big data* e che sarà chiaro soltanto molti anni dopo.

Nel settore informatico l'utilizzo dei *Big data* per finalità commerciali è ancora più diffuso, in particolare nel settore dei videogiochi on-line. Esempio è l'esperienza di Zynga, produttrice di videogames on-line interattivi di grande successo che utilizza i *Big data*, tra i quali in particolare i dati riguardanti le scelte effettive degli utenti all'interno del gioco, per migliorarlo e adattarlo alle necessità degli utenti<sup>76</sup>.

Del secondo tipo di utilizzo a fini commerciali, cioè la vendita direttamente delle informazioni prodotte attraverso l'elaborazione dei *Big data*, si possono riportare alcuni esempi emblematici. Experian, per esempio, con un prodotto chiamato Income Insight, riesce a produrre e vendere delle stime sul livello di reddito delle persone in base alla storia dei loro affidamenti creditizi. Experian ha sviluppato questo indice attraverso un proprio database in cui sono contenute tutte le informazioni sugli affidamenti creditizi, ricostruite in base ai dati fiscali anonimi forniti dallo U.S. Internal Revenue Service<sup>77</sup>. Lo stesso ha fatto un'altra agenzia di informazioni sul credito, come Equifax, che vende due indici (l'“Ability to Pay Index” e un “Discretionary Pay Index”) che dovrebbero prevedere la disponibilità finanziaria delle persone<sup>78</sup>. La stessa Google ha incominciato a vendere informazioni a seguito dell'elaborazione dei *Big data*. Con la collaborazione della banca spagnola BBVA, l'azienda della Silicon Valley vende un servizio di previsioni economico-finanziarie per valutare le prospettive economiche nel settore

---

75 O. ETZIONI, C.A. KNOBLOCK, R. TUCHINDA E A. YATES, *To Buy or Not to Buy: Mining Airfare Data to Minimize Ticket Purchase Price*, SIGKDD '03, 24-25 agosto 2003.

76 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p.194.

77 Per una descrizione più dettagliata dell'esperienza di Experian si può consultare V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 81.

78 Per l'esperienza di Equifax, invece, si v. S. THURM, *Next Frontier in Credit Score: Predicting Personal Behavior*, *Wall Street Journal*, 27 ottobre 2011.

turistico spagnolo<sup>79</sup>. Un altro esempio è Inrix, un'azienda specializzata in analisi del traffico che vende previsioni proprio sui flussi stradali in tempo reale ad aziende che gestiscono flotte commerciali o direttamente ai governi. Inrix offre questo servizio informativo mettendo insieme dati molto differenti tra di loro come i dati provenienti dagli strumenti di geolocalizzazione presenti sulle autovetture, le informazioni che forniscono gli stessi utenti che accedono all'applicazione per la consultazione delle analisi, i dati storici del traffico, i dati sulle condizioni atmosferiche, le informazioni su eventi sportivi o sulle fiere<sup>80</sup>.

Non ci sono soltanto esempi di produzione e messa a disposizione di informazioni per fini commerciali, ma anche esempi di riutilizzo, semplicemente a fini informativi delle informazioni che si possono ricavare dai *Big data*, coerenti con il fenomeno degli *Open data*<sup>81</sup>. Il sito internet FlyOnTime.us, per esempio, utilizzando dati governativi degli USA sul meteo e sui voli, liberamente accessibili tramite le politiche statunitensi sugli *Open data*, segnalano ai visitatori del sito le probabilità di un possibile ritardo dei voli<sup>82</sup>.

Una seconda tipologia di utilizzo dei *Big data* è in funzione di controllo. Molti sono i casi in cui i *Big data* vengono utilizzati a fini preventivi per garantire la stabilità di una struttura, la continuità di un servizio, la sicurezza di un investimento o la legalità delle procedure. Già abbiamo riportato il caso della previsioni sulle esplosioni dei tombini che l'amministrazione comunale di New York ha cercato, con successo, di mettere in atto per tutelare la città dagli effetti nefasti di quelle esplosioni, ma se ne possono riportare altri di casi in cui i *Big data* sono stati utilizzati a fini di controllo preventivo. L'azienda di spedizioni americana UPS conduce delle analisi previsionali sullo stato della sua flotta di autoveicoli per comprendere quando intervenire con le manutenzioni sui veicoli, così come Accenture ha installato una serie di sensori su gran parte della flotta di autobus della città di Saint Louis per lo stesso motivo. Anche la British Petroleum utilizza i *Big data* prodotti dai sensori di rilevazione, per controllare e prevenire guasti all'interno delle proprie raffinerie<sup>83</sup>.

---

79 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 143-147.

80 L'aspetto interessante dell'esperienza di Inrix è che si tratta di una società di intermediazione dei dati. Essa, infatti, riutilizza dati che non possiede e da questi riesce a ricavarne informazioni (e guadagni) rilevanti, che coloro che possiedono i dati non erano stati in grado di sfruttare. Per una descrizione più dettagliata di questa esperienza si v. V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 182-184.

81 Sul fenomeno degli *Open data* e sul forte legame con il fenomeno dei *Big data* si parlerà ampiamente nel capitolo 4.

82 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 160-161.

83 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 84 e 173.

Un altro utilizzo emblematico è quello fatto da ZestFinance, un'azienda fondata dall'ex *chief officer* di Google, che aiuta i finanziatori a decidere se concedere dei prestiti a medio o breve periodo a persone apparentemente non solvibili, affidandosi ad una tecnologia che elabora i *Big data*<sup>84</sup>. Così come alcune grandi compagnie di assicurazioni, come Aviva e AIG, hanno preso seriamente in considerazione la possibilità di individuare i rischi assicurativi sanitari per particolari patologie come diabete, ipertensione arteriosa o depressione, attraverso dati sullo stile di vita delle persone. Utilizzando molte variabili come gli hobby, i siti web visitati, i rapporti sul credito e i dati di marketing, le assicurazioni possono prevedere il rischio sanitario e proporre polizze che tengano conto del rischio previsto<sup>85</sup>.

L'analisi dei *Big data* è stata utilizzata anche per indagare su casi di frode e di corruzione. Il Prof. Steven Levitt, economista dell'Università di Chicago, ha scoperto come ci fossero state, nel tempo, gare del sumo giapponese anomale. Analizzando tutte le gare di undici anni di incontri, oltre 64.000 gare, lo studio ha riscontrato delle anomalie, in quanto, alla fine del campionato, i lottatori di sumo che si trovano più alti in classifica avevano una percentuale di vittorie superiore del 25% rispetto alla media delle vittorie: un'anomalia che ha fatto pensare allo studioso che ci sia la tendenza, alla fine del campionato, a contrattare l'esito di determinate gare<sup>86</sup>. Allo stesso modo l'azienda Xoom, che si occupa di trasferimenti internazionali di denaro, con un'analisi di tutti i dati associati alle sue transazioni, ha scoperto, nel 2011, che erano in corso delle transazioni leggermente superiori alla media in una zona dove era fortemente improbabile che questo accadesse (il New Jersey). Le transazioni, infatti, si scoprirono essere state messe in atto da una banda di criminali<sup>87</sup>.

La terza tipologia di utilizzo dei *Big data* è quella legata al miglioramento della diagnostica medica, al fine di migliorare il giudizio clinico o la conoscenza di un fenomeno sanitario o, comunque, legato alla salute.

Abbiamo già avuto modo di vedere come l'elaborazione dei *Big data* abbia portato a prevedere con probabilità molto alte la possibile diffusione dell'influenza invernale negli USA oppure a individuare delle variabili statistiche che hanno permesso di prevedere le infezioni sui malati prematuri, come è successo nell'Università dell'Ontario. I *Big data* sono stati utilizzati anche per migliorare l'assistenza sanitaria.

---

84 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 70.

85 Su questo punto, molto controverso, si v. L. SCISM – M. MAREMONT, *Insurers Test Data Profiles to Identify Risky Clients*, in *Wall Street Journal*, 19 novembre 2010; ID., *Inside Deloitte's Life-Insurance Assessment Technology*, in *Wall Street Journal*, 19 novembre 2010.

86 M. DUGGAN – S.D. LEVITT, *Winning Isn't Everything: Corruption in Sumo Wrestling*, in *American Economic Review*, 92 (2002), pp. 1594-1605.

87 Esempio riportato in J. ROSENTHAL, *Special Report: International Banking*, in *The Economist*, 19 maggio 2012, pp. 7-8.

Nel 2011 la Fair Isaac Corporation (FICO) ha introdotto il “Medical Adherence Score” per stabilire con quanta probabilità i pazienti assumeranno le medicine loro prescritte dai medici. La FICO mette in relazione tra loro delle variabili che non hanno nessun collegamento apparente, come le informazioni su quanto tempo la persona abita allo stesso indirizzo, da quanto tempo svolge lo stesso lavoro, se possiede un'auto, ma i risultati sono stati molto utili all'assistenza sanitaria statunitense nel sollecitare, in maniera mirata, l'assunzione dei medicinali<sup>88</sup>. Vale la pena riportare altri due esempi. Il primo riguarda l'utilizzo dei Big data per migliorare l'organizzazione sanitaria. Il MedStar Hospital Center di Washington utilizza il software Amalga della Microsoft per cercare delle soluzioni per ridurre i ricoveri reiterati e le infezioni, che costituiscono le voci di spesa più elevate per la sanità statunitense, attraverso l'utilizzo dei dati storici anonimi sui pazienti in suo possesso, come dati anagrafici, esami clinici, diagnosi, terapie etc.. Queste correlazioni hanno fatto emergere un elenco di variabili che facevano aumentare le probabilità di ricovero del paziente entro un mese dalle dimissioni, tra cui alcune abbastanza prevedibili come determinate malattie difficili da curare in via definitiva, altre meno intuitive come lo stato psicologico del paziente<sup>89</sup>.

Anche l'amministrazione statunitense sta iniziando ad utilizzare i *Big data*. Il Dipartimento *Veteran Affairs*, per esempio, utilizza i *Big data* per monitorare lo stato di salute dei veterani e per migliorare la gestione dei dati e l'integrazione delle banche dati<sup>90</sup>. Allo stesso modo il Dipartimento Salute e Servizi umani (HHS), il Dipartimento che si occupa dei famosi programmi di assistenza sanitaria Medicare e Medicaid, utilizza i *Big data* per una serie di finalità. Innanzitutto sta tentando di integrare il sistema sanitario federale dei dati con il sistema sanitario degli Stati, attraverso l'integrazione delle banche dati.<sup>91</sup>

Nel settore sanitario è utile anche riportare come il fenomeno dei *Big data* abbia notevolmente influito sulla capacità del settore di reperire nuovi e ulteriori dati direttamente dai paziente, attraverso la c.d. medicina partecipata, in cui il paziente fornisce preziose informazioni, direttamente con il suo corpo, che prima erano difficili da reperire se non nelle strutture mediche. Un esempio di datizzazione del corpo è il

---

88 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 81.

89 Informazioni più approfondite possono essere consultate su V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 173-174, il quale richiama un comunicato stampa della Microsoft del 26 luglio 2006 *Microsoft Expands Presence in Healthcare IT Industry with Acquisition of Health Intelligence Software Azyxxi*.

90 Executive Office of the President - Office of Science and Technology Policy, *Fact Sheet: Big Data Across the Federal Government*, Washington, marzo 2012. Questo documento, che citeremo più volte in questo lavoro, contiene una prima ricognizione dell'utilizzo dei *Big data* all'interno dell'amministrazione americana.

91 Ibidem.

“*quantified self*”, un modo per datizzare tutto gli elementi biometrici del proprio corpo in tempo reale<sup>92</sup>. Un altro è la start-up statunitense Zeo che attraverso l'elaborazione dei dati trasmessi dagli smartphone ha già trovato il modo di datizzare le dinamiche del sonno<sup>93</sup>. Anche il National Institut of Health utilizza i *Big data* prodotti dall'*imaging* medico, in particolare *neuroimaging* per la cura dell'Alzheimer<sup>94</sup>.

Ma gli esempi di medicina partecipata e datizzazione del corpo del paziente sono tantissimi e tutti rilevanti per molteplici aspetti, dalla diagnostica al giudizio medico, dall'organizzazione sanitaria alla tutela della riservatezza dei dati.

Un'altra tipologia di utilizzo dei *Big data* è in funzione di sicurezza, intesa in via generale: nel settore della difesa, per il controllo del territorio e per il contrasto alla criminalità e al terrorismo.

Nel settore della difesa è emblematico come il Dipartimento della Difesa dell'amministrazione statunitense utilizzi i *Big data* per le operazioni militari, attraverso l'elaborazione dei dati e delle immagini, attraverso l'interazione uomo-macchina, anche per azioni di difesa mirate (con i famosi attacchi a distanza) e attraverso lo sviluppo di sistemi di protezione informatica che proteggono i sistemi di difesa da attacchi cybernetici<sup>95</sup>. Anche il Dipartimento della *Homeland Security* (DHS) si sta preparando ad utilizzare in maniera più intensa i *Big data*. Il Centro di Eccellenza sulla visualizzazione e analisi dei dati (CVADA), parte del DHS, in collaborazione con alcuni ricercatori della Rutgers University e della Purdue University, conduce attività di ricerca attraverso un programma, First Responder, che utilizza i *Big data* per affrontare problemi che vanno dalle catastrofi artificiale o naturali per atti terroristici al controllo delle frontiere passando per la prevenzione contro le minacce informatiche<sup>96</sup>.

Anche il controllo del territorio, inteso come prevenzione dei disastri naturali o informazioni in via preventiva dei fenomeni naturali, è stato un terreno in cui i *Big data* hanno avuto sbocchi di utilizzo. I *Big data* prodotti dai social network e dagli strumenti di geolocalizzazione sono stati determinanti per il soccorso dopo l'uragano Irene, negli USA, così come il sito FloridaDisaster.org, che utilizza basi di dati differenti, cioè i dati del web, i dati geografici, i dati sul meteo per fare informazione sugli eventi naturali di una certa rilevanza o sulle catastrofi naturali<sup>97</sup>.

92 Un esempio riportato in *Counting every moment*, in *The Economist*, 3 marzo 2012.

93 Questo e numerosi altri esempi di datizzazione degli elementi biometrici in V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 132-133.

94 Executive Office of the President - Office of Science and Technology Policy, *Fact Sheet: Big Data Across the Federal Government*, Washington, marzo 2012.

95 Ibidem.

96 Ibidem.

97 A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, Maggioli, Bologna, 2013, p. 31.

I *Big data* sono stati utilizzati anche per il controllo e la sorveglianza del territorio in funzione anti-terroristica. Sono stati più volte documentati vari programmi per la sorveglianza preventiva: il Washington Post nel 2010 ha informato i cittadini che la NSA aveva raccolto miliardi di dati informatici e personali dei cittadini negli Stati Uniti<sup>98</sup>. Una sorveglianza sistemica che è cresciuta e cresce ogni qualvolta ci sono attacchi terroristici e che molti studiosi hanno documentato e analizzato in più occasioni<sup>99</sup>, dimostrando come la raccolta di enormi quantità di dati sulle abitudini, sulle conversazioni e sui movimenti dei cittadini siano fondamentali allo scopo. Ci sono stati anche progetti di ricerca patrocinati dal Dipartimento della Homeland Security come quello denominato FAST (Future Attribute Screening Technology) che hanno tentato di identificare i potenziali terroristi monitorando i segni vitali, il linguaggio non verbale, gli aspetti fisiologici per rivelare l'intento criminoso con un'accuratezza del 70%<sup>100</sup>.

Sono sempre più diffusi nelle città statunitensi programmi di sorveglianza preventiva che utilizzano le analisi dei *Big data* per selezionare le strade, i gruppi sociali o gli individui da tenere sotto osservazione in quanto probabilmente sono luoghi ad alta intensità criminale o probabilmente hanno un'alta propensione alla criminalità. Emblematico di questo approccio è il programma Blue CRUSH (Crime Reduction Utilizing Statistical History), utilizzato nella città di Memphis, in Tennessee. Questo programma aiuta i funzionari e gli agenti di polizia ad individuare le aree in cui potrebbe essere probabile che avvengano furti, rapine o omicidi, dando informazioni precise su dove (determinati isolati) e quando (alcune ore di un determinato giorno della settimana) questi reati potrebbero essere commessi<sup>101</sup>. Allo stesso modo a Richmond, in Virginia, la polizia correla i dati sulla criminalità con altri *dataset* - come le date di pagamento degli stipendi, dei concerti o delle fiere - per capire quando e dove probabilmente potrebbero esserci dei picchi di criminalità. La polizia aveva sempre sospettato, infatti, che il livello di allerta dovesse essere maggiore dopo le fiere sulle armi e l'analisi dei dati ha dimostrato che i sospetti erano fondati<sup>102</sup>.

Un professore di statistica e criminologia alla Università della Pennsylvania Richard Berk, ancora, ha affermato che il suo metodo di ricerca, basato sull'elaborazione dei *Big*

---

98 Per queste notizie si v. D. PRIEST – W. ARKIN, *A Hidden World, Growing Beyond Control*, in *The Washington Post*, 19 luglio 2010; J. GONZALEZ, *Whistleblower: The NSA Is Lying – U.S. Government Has Copies of Most of Your Emails*, in *Democracy Now*, 20 aprile 2012; W. BINNEY, *Sworn Declaration in the Case of Jewel vs. NSA*, 2 luglio 2012.

99 Per tutti si v. diffusamente D. LYON, *Massima Sicurezza. Sorveglianza e "guerra al terrorismo"*, trad. it. E. Greblo, Milano, 2005.

100 S. WEINBERGER, *Terrorist «Pre-Crime» Detector Field Tested in United States*, in *Nature*, 27 maggio 2011. Sui casi di falsi positivi si v. anche A. FURNAS, *Homeland Security's «Pre-Crime» Screening Will Never Work*, in *The Atlantic Online*, 17 aprile 2012.

101 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 214.

102 Ibidem.

*data*, fosse in grado di prevedere, con una probabilità minima del 75%, se un detenuto, una volta uscito dagli istituti di detenzione, potesse essere coinvolto in un reato, attraverso l'utilizzo di alcune variabili, tra cui la data del primo reato, il motivo della detenzione e i dati anagrafici<sup>103</sup>.

L'ultima tipologia di utilizzo è quella in funzione conoscitiva. Questa è certamente l'utilizzo più immediato dei *Big data* e l'utilizzo che in fondo caratterizza tutta la loro logica. I *Big data* vengono utilizzati principalmente per conoscere (per poi comprendere) la realtà che circonda coloro che devono prendere delle decisioni. Essi hanno, quindi, innanzitutto una funzione conoscitiva strumentale al processo decisionale. Tutti gli esempi che abbiamo elencato fino ad ora in questo lavoro sono caratterizzati dal fatto che l'utilizzo dei *Big data* ha prodotto un avanzamento conoscitivo utile soprattutto per prendere delle decisioni concrete. Si pensi ai casi che abbiamo riportato nell'ambito sanitario o nel governo del territorio. Un altro caso che vale la pena riportare è quello che ha visto protagonisti la casa automobilistica Honda, l'IBM e la compagnia californiana Pacific Gas and Electric Company. Le tre aziende hanno avviato un progetto di ricerca nel 2012 per raccogliere tutte le informazioni possibili per impiantare in California un sistema di rifornimento per le auto elettriche e per comprendere quale impatto avrebbe avuto sulla rete elettrica. Il sistema previsionale che è scaturito da questa collaborazione - utilizzando dati storici e presenti della rete elettrica, il livello di carica della batteria, l'ubicazione della vettura, l'ora e i punti di ricarica disponibili - non solo è riuscito a determinare i tempi e i luoghi ideali per mettere a punto il sistema, ma, con ulteriori dati provenienti dal meteo e dai prezzi delle varie stazioni, potrebbe anche permettere a coloro che gestiscono il servizio di utilizzare i dati e le informazioni per migliorarlo ulteriormente<sup>104</sup>.

In secondo luogo, oltre ad una funzione conoscitiva strumentale al processo decisionale, i *Big data* sono stati utilizzati anche per un più generale avanzamento della conoscenza dei fenomeni, non finalizzata strettamente alla necessità di dover prendere delle decisioni. Si pensi alle ricerche che hanno permesso, grazie all'utilizzo di correlazioni non lineari, di confermare la correlazione tra felicità e reddito. Un legame che tutta la teoria economica e politologica è abituata a considerare per acquisito, a cui i *Big data* sono riusciti a dare una base scientifica e a sottolineare la sua complessità. Le correlazioni non lineari, a differenza di quelle tradizionali, hanno rilevato come al variare del reddito, oltre una certa soglia reddituale, l'effetto psicologico positivo non cresceva più: per aumentare la soddisfazione sociale bisognava, quindi, ridurre le disuguaglianze e fare crescere il reddito delle classi sociali più povere<sup>105</sup>. Oppure si

---

103 Ibidem.

104 C. LUTHY, *Guest Perspective: IBM Working with PG&E to Maximize the EV Potential*, in *PGE Currents*, 13 aprile 2012.

105 R. INGLEHART – H.D. KLINGEMANN, *Genes, Culture and Happiness*, MIT Press, Cambridge, 2000.

pensi a come dei ricercatori della Harvard e del MIT, utilizzando i *Big data*, siano stati capaci di scoprire come la disparità dei livelli di immunizzazione contro la rosolia nella popolazione statunitense, correlata alla spesa sanitaria individuale, diventava più bassa se si abbassava la spesa sanitaria e viceversa. Questo dato dimostrava come la popolazione più agiata diffidasse del vaccino contro la rosolia a differenza della popolazione più povera<sup>106</sup>. È utile riportare, inoltre, l'esperimento di due economisti del MIT, Alberto Cavallo e Roberto Rigobon, che hanno elaborato un apposito software che, dopo aver raccolto milioni di dati sui prezzi dei prodotti venduti negli USA, ha rilevato da subito un'oscillazione deflazionistica dopo il fallimento della Lehman Brothers nel 2008 qualche mese prima dei dati ufficiali del Consumer Price Index forniti dallo U. S. Bureau of Labor Statistics<sup>107</sup>.

La stessa amministrazione americana, infine, in particolare gli uffici e i laboratori del Dipartimento dell'Energia, la NASA e vari Istituti nazionali (come il General Medical Science), utilizza i *Big data* a fini prettamente conoscitivi per migliorare le rilevazioni, le ricerche o semplicemente per coadiuvare la ricerca attraverso l'integrazione dei *dataset*<sup>108</sup>.

Questa breve elencazione delle tipologie di utilizzo dei *Big data* mette in evidenza come non si sta trattando di un fenomeno marginale e aleatorio, ma, e questo in maniera sempre più evidente, di un fenomeno che sta assumendo una concretezza e una diffusione crescenti. Il suo impatto sulla società, sulle sue regole e sui singoli individui può essere, come abbiamo potute vedere, rilevante e molte volte anche pervasivo. L'utilizzo dei *Big data* ha certamente delle enormi potenzialità, soprattutto se prendiamo in considerazione gli effetti, in gran parte positivi, che esso ha avuto sulla cultura dei dati e del loro utilizzo, la quale ci costringe a mettere in dubbio impostazioni culturali molto radicate, che caratterizzano profondamente il nostro modo di osservare, analizzare ed modificare la realtà che ci circonda. I fenomeni conoscitivi come la datizzazione della realtà, l'utilizzo di correlazioni e analisi predittive nell'elaborazione dei dati, la cultura dell'estrema valorizzazione del dato, hanno fatto emergere tutte le potenzialità conoscitive dei *Big data*.

Abbiamo avuto modo di elencare una serie di esperienze concrete che hanno evidenziato come i *Big data* sono capaci di modificare profondamente i processi decisionali, pubblici e privati, in quanto capaci di fornire conoscenze approfondite e punti di vista originali sui fenomeni scientifici, sociali, economici e culturali in atto;

---

106 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 62.

107 A. Lowrey, *Economists's Programs are beating U.S. at Tracking Inflation*, in *Washington Post*, 25 dicembre 2010.

108 Executive Office of the President - Office of Science and Technology Policy, *Fact Sheet: Big Data Across the Federal Government*, Washington, marzo 2012.

sono capaci di modificare profondamente la produttività delle imprese e di interi settori economici creando nuove competenze, nuovi spazi di mercato, un nuovo paradigma concorrenziale, tutti aspetti legati alla capacità di sapere utilizzare i dati.

La centralità del dato, inoltre, apre prospettive inedite, che alcuni studiosi hanno messo bene in evidenza<sup>109</sup>: in futuro i dati potrebbero costituire degli *asset* talmente tanto importanti da essere contabilizzati nei bilanci aziendali e fare parte dei c.d. “*asset* intangibili” delle aziende, paragonabili al *brand*, alla strategia e alle potenzialità produttive, in sintesi, tutto ciò che non rientra nella contabilità finanziaria dell'azienda.

Ma la centralità del dato potrebbe anche mettere al centro la persona, in maniera del tutto inedita. Se, come abbiamo potuto vedere dagli esempi fatti nel corso di questo scritto, l'individuo è una delle maggiori fonti di produzione di *Big data*, in futuro potrebbe cambiare il rapporto tra la persona e i propri dati: l'individuo potrebbe avere un ruolo centrale nella gestione dei propri dati mai avuto precedentemente; potrebbe utilizzare i dati, da solo o in forme collettive, in modo tale da valorizzare le sue prerogative singole o collettive.

Il fenomeno dei *Big data*, però, non ha solo delle intrinseche potenzialità, ma anche limiti, alcuni di una certa rilevanza, non necessariamente legati al loro cattivo utilizzo, ma al loro utilizzo *tout court*. Dagli esempi concreti appena elencati si possono individuare alcuni grandi limiti dei *Big data*, che producono anche dei rischi rilevanti e che costituiscono due facce della stessa medaglia.

Il primo limite riguarda la pervasività dello strumento, che produce, innanzitutto, effetti negativi sulla *privacy* degli individui. Si pensi alle strategie di marketing o all'utilizzo dei *Big data* per la sorveglianza, che, non solo ripropongono i rischi per la *privacy* che erano legati alla questione del *profiling*, ma ne potenziano enormemente gli effetti. In secondo luogo la pervasività degli strumenti di analisi dei *Big data* sta creando un rischio per l'individuo anche a causa delle correlazioni frutto dell'utilizzo di quegli strumenti: gli esempi hanno evidenziato come gli algoritmi sono capaci di dirci quali sono le probabilità di subire un attacco di cuore, con delle pesanti ricadute sui premi assicurativi sanitari, di non riuscire a rimborsare un mutuo ipotecario, con la possibilità che venga negato il finanziamento, o di commettere un crimine, con la possibilità di essere strettamente sorvegliato.

Il secondo grande limite dei *Big data*, che rappresenta l'altro lato della medaglia, coinvolge direttamente coloro che utilizzano i dati: alcuni studiosi, infatti, hanno parlato del rischio di una vera e propria “dittatura dei dati”<sup>110</sup>. Possedere una quantità enorme di dati e avere tutte le capacità tecniche e culturali per raccoglierne ancora e in merito a

---

109 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 161-165.

tutti i fenomeni esistenti, umani e naturali, può portare qualsiasi decisore, pubblico o privato che sia, a credere di potere risolvere qualsiasi problema, di fare fronte a qualsiasi necessità attraverso l'utilizzo dei dati. Fidarsi ciecamente dei dati, lasciarsi governare da loro, essendo totalmente offuscati dai loro limiti e prendere delle decisioni unicamente sulla base delle loro elaborazioni, poter credere di ricorrere sempre all'elaborazione dei dati per risolvere qualsiasi criticità o per rispondere a qualsiasi bisogno, rischiano di appiattire la decisione su logiche troppo quantitative e di renderla, anziché più completa e adeguata ai problemi e alle situazioni che è necessario affrontare, fortemente parziale e largamente incompleta. I dati per quanto possano essere di ottima qualità, analizzati con una procedura corretta e usati nella maniera più corretta (e molte volte non lo sono), non possono essere l'unico metro di valutazione per risolvere problemi per lo più complessi, all'interno dei quali la persona, i suoi comportamenti e le sue dinamiche sociali possono essere datizzate, ma mai pienamente interpretabili dai dati.

---

110 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 220 e ss..

## Capitolo 2

### La dimensione giuridica dei *Big data*

Sommario: 1. Il rilievo giuridico dei *Big data*.; 2. *Big data*, mercati e concorrenza.; 3. *Big data* e il profilo individuale della tutela dei dati personali.; 4. I *Big data* e il profilo collettivo della tutela dei dati individuali.; 5. *Big data*, sorveglianza e prevenzione ai fini di sicurezza e controllo.; 6. *Big data*, potere e asimmetrie informative.; 7. I *Big data* e la regolazione pubblica.; 8. I *Big data* e le amministrazioni pubbliche.;

#### 1. Il rilievo giuridico dei *Big data*.

Il fenomeno dei *Big data* è un fenomeno complesso che sta influenzando pesantemente sulla gestione dei dati e sulla produzione di conoscenza pubblica e privata. Un fenomeno con una dimensione ambivalente: una quantitativa e l'altra prettamente qualitativa. Le caratteristiche dei *Big data*, la loro ragguardevole quantità, tendenzialmente mai vista prima e in costante espansione, la loro eterogeneità nelle forme e nelle fonti di produzione, la loro dimensione quantitativa appunto, hanno reso obsolete gran parte delle tecnologie di conservazione ed elaborazione dei dati e hanno permesso lo sviluppo della prima dimensione qualitativa del fenomeno, cioè lo sviluppo di nuove tecnologie in grado di sfruttare pienamente le loro potenzialità.

Riuscire a datizzare la realtà che ci circonda, poter osservare i fenomeni sotto altri punti di vista o, addirittura prevederne le loro evoluzioni, valorizzare dati che prima non si sarebbero mai riusciti a reperire (si pensi ai dati residui delle navigazioni su internet) o che non era necessario neanche utilizzare (si pensi ai dati di cattiva qualità), elementi che caratterizzano la seconda dimensione qualitativa dei *Big data*, hanno messo pesantemente in discussione molte delle categorie culturali che hanno caratterizzato la gestione e l'elaborazione dei dati fino ad oggi. La logica degli *Small data*, costituita da una costante attenzione alla qualità e alla correttezza dei dati che si conservano e che si utilizzano, dalla necessità di darsi regole tassonomiche strutturate, dalla volontà di volere spiegare i fenomeni che si analizzano, dalla tendenza ad utilizzare i dati fondamentalmente per la finalità per cui sono stati raccolti, sta lasciando progressivamente il passo alla logica dei *Big data*, in cui la qualità e le regole tassonomiche non sono poi così importanti perché è primario avere tutti i dati a disposizione, in cui la volontà di osservare la realtà è più utile della volontà di spiegare la realtà, in cui i dati sono una risorsa da utilizzare più e più volte anziché una volta sola.

Un fenomeno che, come abbiamo visto nel capitolo precedente, si sta diffondendo esponenzialmente. Le variegate tipologie di utilizzo ne sono una prova evidente.

Abbiamo visto che i *Big data* vengono utilizzati per finalità commerciali. Le aziende, infatti, utilizzano i dati per migliorare la loro produttività, per migliorare i servizi esistenti, per creare nuovi servizi, ma anche per elaborare i dati e vendere le informazioni che si sono prodotte. I *Big data* vengono utilizzati anche in funzione di controllo per garantire la stabilità di una struttura o di un'infrastruttura, la continuità di un servizio, la sicurezza di un investimento o, in alcuni casi, anche per garantire il rispetto della legalità delle procedure, anche a finalità antifrode o anticorruzione. Particolarmente interessanti sono stati gli utilizzi che sono stati fatti nel settore sanitario, per migliorare la diagnostica medica, il giudizio clinico o la conoscenza di un fenomeno sanitario o, comunque, legato alla salute. Rilevante è stato anche l'utilizzo dei *Big data* in funzione di sicurezza, nel settore della difesa, ma anche per il controllo del territorio e per il contrasto alla criminalità e al terrorismo. In ultimo i *Big data* sono stati utilizzati ampiamente in funzione conoscitiva, intesa sia come supporto al processo decisionale, sia come avanzamento della conoscenza nei più disparati settori scientifici.

Le tipologie di utilizzo ci dimostrano come il fenomeno abbia già assunto una sua concretezza, la quale, di fatto, ha permesso ad alcuni studiosi di sistematizzare le caratteristiche e le peculiarità del fenomeno in maniera completa<sup>111</sup>, e ha anche fatto emergere i primi rilievi giuridici del fenomeno. Quando nel capitolo precedente abbiamo cominciato ad elencare le potenzialità e i limiti dei *Big data*, alcuni aspetti di indubbio rilievo giuridico sono emersi. Si pensi all'evoluzione delle dinamiche concorrenziali all'interno dei mercati, alla tutela dei dati personali, ai diritti dei cittadini. Aspetti che stanno cominciando ad essere analizzati, alcuni in maniera più approfondita altri meno, dalla letteratura scientifica e, soprattutto, da molti rapporti commissionati e prodotti da organizzazioni pubbliche e private a livello internazionale.

In questo capitolo, infatti, cercheremo - in maniera più completa possibile, ma senza pretendere di essere esaustivi - di elencare e cercare di analizzare quale rilievo giuridico sta avendo concretamente il fenomeno dei *Big data*, appunto richiamando, per quanto sarà possibile, studi o letteratura scientifica che hanno cominciato a sottolineare l'impatto e la rilevanza del fenomeno in termini giuridici.

---

111 Il volume di V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., resta ancora determinante per la sua capacità di analizzare il fenomeno in maniera complessiva.

## 2. *Big data, mercati e concorrenza.*

Essendo un fenomeno nato principalmente nel settore privato, i *Big data* hanno avuto degli effetti rilevanti in molti settori economici<sup>112</sup>. Il loro ingresso nelle dinamiche di mercato ha avuto come principale ricaduta giuridica, trasversale a tutti i settori, ma particolarmente evidente all'interno dei mercati digitali e dei servizi, quello di mettere in crisi alcuni aspetti, anche fondamentali, del diritto antitrust. L'influenza del fenomeno dei *Big data* sulle dinamiche di mercato attiene fondamentalmente a due dimensioni: quella più generale, riguardante il possesso (e il successivo utilizzo) dei dati e le sue ricadute sui rapporti di forza tra le aziende che si confrontano sul mercato e quella inerente l'utilizzo dei *Big data* e gli effetti che esso produce sui consumatori.

Nonostante il fenomeno sia in espansione e i costi delle tecnologie utili per gestire i *Big data* stiano diminuendo, sono ancora poche le aziende che riescono a produrre e ad utilizzare a pieno le potenzialità dei *Big data* nelle loro attività economiche e, in particolare, commerciali. Il fenomeno dei *Big data* ha creato, nel tempo, delle profonde disparità all'interno di alcuni mercati, tra le aziende che possiedono i dati e producono le informazioni, in particolare informazioni in merito agli utenti e ai consumatori, e coloro che ne possiedono di meno o non ne possiedono affatto e, di conseguenza, hanno meno informazioni sui consumatori e sugli utenti. Le situazioni di monopolio che si vengono a creare, infatti, sono fortemente legate alla produzione e alla gestione di informazioni sui consumatori e alla capacità delle aziende di “conoscere” il più possibile le preferenze e i comportamenti dei consumatori a cui ci si vogliono rivolgere. Le vere disparità di mercato, quindi, sono quelle che intercorrono tra coloro che “conoscono”, attraverso le elaborazioni dei *Big data*, i consumatori e coloro che “conoscono” poco i consumatori, in quanto non possiedono i *Big data* o non hanno le tecnologie adatte per elaborarli. La capacità di ricostruire il profilo dei clienti o dei potenziali clienti che l'elaborazione dei *Big data* permette di fare in maniera più semplice rispetto al passato, facilita il miglioramento dei servizi offerti o le capacità di pubblicizzare e vendere i propri prodotti a scapito della concorrenza, alcune volte, in maniera difficilmente reversibile. Un divario che stabilizza un vantaggio competitivo evidente a favore delle aziende che “conoscono” i consumatori e gli utenti, che mette loro non solo in una condizione di monopolio, in quanto possessori di dati e di informazioni che le aziende

---

<sup>112</sup> Si pensi all'influenza che le *Smart grid*, cioè le reti energetiche intelligenti utilizzate per l'efficientamento energetico, stanno avendo nel mercato dell'energia. Queste reti vengono utilizzate sempre più spesso non solo per raccogliere dati sui consumatori, anche personali, ai fini di migliorare le performance energetiche delle reti, ma sempre più spesso per scopi prettamente commerciali. Per un maggiore approfondimento con una particolare attenzione alle dinamiche europee e nazionali si v. F. GIGLIONI, *La sfida dell'innovazione sulla regolazione pubblica. il caso delle "smart grid"*, in *Mumus*, 2013, fasc. 3, pag. 463 – 496.

concorrenti non hanno, ma anche di potere mantenere quel monopolio proprio perché, avere una quantità di dati e di informazioni così grande e in continuo aggiornamento, anche grazie al contributo degli stessi consumatori, dà la possibilità a quelle aziende di raggiungere un livello competitivo difficilmente superabile<sup>113</sup>.

Questi aspetti hanno cominciato a fare emergere degli interrogativi rilevanti sulla capacità della disciplina della concorrenza di regolare le nuove dinamiche del mercato e frenare i monopoli che si stanno consolidando, in particolare ci si interroga sulla capacità del diritto antitrust di fare fronte ad un fenomeno fondamentalmente nuovo: al protagonismo, molte volte inconsapevole, dei consumatori nelle dinamiche concorrenziali tra le aziende. Questi, infatti, fornendo continuamente e volontariamente dati personali, sulle proprie abitudini e sui propri comportamenti, contribuiscono per primi a creare e a rafforzare le posizioni dominanti nel mercato. In una situazione come quella appena descritta, ci si è domandati, per esempio, se definizioni come quella di mercato rilevante - contenuta nella disciplina europea e principio cardine della disciplina antitrust per la valutazione dell'esistenza di pratiche restrittive o di abuso della posizione dominante nel mercato unico europeo<sup>114</sup> - possano essere ancora parametri guida, ad esempio in mercati come quello digitale, per ricostruire le caratteristiche del mercato e per individuare probabili posizioni dominanti<sup>115</sup>.

La capacità di raccogliere tutte le informazioni possibili sui consumatori o sugli utenti per migliorare e personalizzare offerte commerciali e servizi sta, inoltre, incentivando le aziende che possiedono i *Big data* ad utilizzare sempre più spesso strategie di prezzo

---

113 Mette in evidenza questi aspetti MANTELETO A., *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, in *Il Diritto dell'informazione e dell'informatica*, 2012, fasc. 1, pag. 135 – 144; D. BOLLIER, *The promise and peril of Big data*, Washington, *The Aspen Institute*, 2010; MCKINSEY, *Big data: the next frontier for innovation, competition and productivity*, 2011; European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.

114 La definizione di mercato rilevante e la procedura per la sua identificazione è stata descritta nel dettaglio all'interno della *Comunicazione della Commissione sulla definizione di mercato rilevante ai fini dell'applicazione del diritto comunitario in materia di concorrenza* [Gazzetta ufficiale C 372 del 9.12.1997]. Definire il mercato rilevante significa stabilire la portata delle regole sulla concorrenza rispetto alle pratiche restrittive e all'abuso di posizione dominante, disciplinate dal Regolamento (CE) n. 1/2003, e la portata dei regolamenti sulle fusioni, regolate dal Regolamento (CE) n. 139/2004. La Commissione opera una valutazione della sostituibilità sul versante della domanda e sul versante dell'offerta e sulle effettive condizioni in cui operano le imprese. Per mercati come quelli digitali, una valutazione di questo genere non basta per interpretarne a fondo le specificità in quanto i prodotti e gli operatori sono caratterizzati diversamente rispetto ai mercati tradizionali.

115 In questo senso e con riferimenti anche all'esperienza statunitense si v. l'intervento del Prof. A. ODLYZKO, *Privacy, price discrimination and seeds of capitalism's destruction*, 28 ottobre 2015, Camera dei deputati organizzato dall'Intergruppo parlamentare sull'innovazione tecnologica. Anche European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, cit.

differenziate per fasce di utenti e consumatori. Le strategie di prezzo differenziate, come tutti i fenomeni legati all'utilizzo dei *Big data*, non sono ovviamente una pratica nuova, ma sono diventate molto più invasive e, se vogliamo, molto più efficaci nel targetizzare e costruire fasce di utenti e consumatori sempre più omogenei e dettagliate. Le aziende, infatti, fruttano in maniera sempre più intensa i *Big data* per creare profili di utenti e di consumatori sempre più accurati in maniera tale da proporre a ciascuno un prezzo diverso del loro prodotto o servizio, un prezzo che sarà tanto più alto quanto maggiore sarà la disponibilità del cliente a spendere per quel prodotto. In questo caso, quindi, i *Big data* non solo vengono utilizzati per profilare i clienti e gli utenti, ma anche per calcolare, attraverso algoritmi complessi e sempre più precisi, le previsioni di disponibilità della spesa e, sulla base di queste previsioni, dividere gli utenti e i consumatori per fasce di prezzo.

La strategia di differenziazione dei prezzi se, da un lato, possono favorire la produttività dell'azienda che la utilizza e anche accrescere la dinamicità del mercato, dall'altro, un loro utilizzo eccessivo può ledere le tradizionali dinamiche di domanda e offerta su cui la concorrenza si basa e può creare delle forti barriere di ingresso per i nuovi operatori del mercato, i quali non potendo utilizzare queste strategie, non riescono a incrinare le posizioni dei giganti del settore<sup>116</sup>.

Il tema della *price discrimination*, però, coinvolge, in maniera molto più rilevante, un altro profilo, che attiene alla tutela del consumatore ed è una diretta conseguenza del meccanismo sottostante la differenziazione dei prezzi, ed è quello inerente la possibile discriminazione a cui i consumatori sono sottoposti. Le fasce di prezzo, infatti, vengono stabilite essenzialmente da una inferenza statistica e la stessa operazione si occupa di inserire il singolo consumatore all'interno di una determinata fascia e rischia di influire pesantemente su aspetti rilevanti della vita del singolo, come ad esempio sull'esclusione sociale<sup>117</sup>.

Le dinamiche legate alla *price discrimination* non sono le uniche che hanno fatto emergere nuove criticità per la disciplina sulla tutela del consumatore. È emerso, anche in rilevanti documenti prodotti dall'amministrazione americana, come l'opacità sulla raccolta delle informazioni e sull'utilizzo dei dati dei consumatori, soprattutto quelli personali, a fini pubblicitari e di marketing e la mancanza di chiarezza sul funzionamento degli algoritmi che elaborano questi dati, possano produrre forti asimmetrie informative tra le aziende e i consumatori. Il consumatore medio, di solito, non solo si trova nella condizione di non conoscere chi detiene i propri dati, quali sono i

---

116 Ibidem.

117 Sugli effetti discriminatori delle strategie di prezzo differenziate interessante è il documento dell'*Executive Office of the President of the U.S.A.*, *Big data and differential pricing*, Washington, Febbraio 2015 e la bibliografia in esso contenuta.

dati personali che sono stati raccolti e quali dati personali vengono utilizzati, ma può essere vittima di una visione limitata del mercato e dei prodotti *on-line*, in quanto queste informazioni sono veicolate da algoritmi che vengono influenzati dalle preferenze e dalle scelte del consumatore stesso. Queste dinamiche rischiano non solo di ledere il diritto dei consumatori alla trasparenza e alla pubblicità delle informazioni sui servizi offerti e sulle metodologie utilizzate per pubblicizzarle, in particolare quando si tratta di pubblicità *on-line*<sup>118</sup>, ma rischiano di intrappolare il consumatore in una sorta di effetto *lock-in*<sup>119</sup>.

Un ultimo aspetto a cui vale la pena accennare e che attiene agli effetti prodotti dai *Big data* sulle dinamiche concorrenziali è quello legato alla disciplina del diritto d'autore, in particolare alla disciplina che tutela la creazione e la formazione delle banche dati. L'utilizzo esclusivo di alcune informazioni e la tutela delle banche dati che le contengono diventano, proprio per gli aspetti che abbiamo sottolineato in precedenza, determinanti e un bene da proteggere. Alcuni studiosi, però, hanno sottolineato come la costruzione delle banche dati che contengono *Big data* possano essere tutelate solo marginalmente dall'attuale disciplina<sup>120</sup>. Il diritto d'autore tutela principalmente «le banche dati che per la scelta o per la disposizione del materiale costituiscono una creazione intellettuale dell'autore»<sup>121</sup> che, secondo una unanime interpretazione dottrinale<sup>122</sup>, hanno come elemento fondamentale il fatto che, rispecchiando fondamentalmente la personalità dell'autore, sono considerate come sue vere e proprie opere di ingegno dello stesso. La legge n. 633/1941 tutela anche le c.d. banche dati non creative all'art. 102-bis, banche dati su cui il costituente, cioè colui che «effettua

---

118 In questo senso vanno due importanti documenti pubblicati dalla Casa Bianca, in cui l'amministrazione americana comincia a pensare a politiche di maggiore tutela del consumatore nei confronti del fenomeno dei *Big data*. Il primo è dell'*Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014 e il secondo è *Executive Office of the President of the U.S.A., Big data and differential pricing*, Washington, Febbraio 2015. Nell'Unione europea è possibile reperire delle prime impressioni sul punto in European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.

119 Per effetto *lock-in* sui consumatori si intende qualsiasi strumento che permette alle aziende di creare delle barriere all'entrata nei confronti di potenziali concorrenti e di acquisire un forte potere di mercato. In questo caso utilizzare le informazioni sulle preferenze e sulle attività dei consumatori per direzionare e selezionare i prodotti da comprare può generare un pericoloso effetto *lock-in*. In questo senso European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.

120 Si v. M. BOGNI – A. DEFANT, *Big data: diritti IP e problemi della privacy*, in *Dir. Ind.*, fasc. 2, 2015;

121 art. 2 della legge n. 633/1941.

122 Per una panoramica generale sulle disposizioni della legge n. 633/1941 sulle banche dati *sui generis* così come adattata alla direttiva 96/9/CE, che disciplina il diritto d'autore a livello europeo, si v. G. GUGLIELMETTI, *La tutela della banche dati con diritto sui generis nella direttiva 96/9/CE*, in *Contr. e impr. Europa*, 1997, p. 181.

investimenti rilevanti per la costruzione di una banca dati o per la sua verifica o la sua presentazione, impegnando a tal fine, mezzi finanziari, tempo o lavoro». Queste disposizioni, secondo le prime interpretazioni e analisi del fenomeno, non sono in grado di interpretare i cambiamenti e le specialità connessi alle banche dati contenenti i *Big data* in quanto rientrerebbero soltanto nella seconda definizione, quindi protette soltanto da un diritto connesso in capo al costituente (quello di impedire a chiunque di fare operazioni di estrazione o di impiego della totalità o di una parte di dati) e che ha un contenuto del tutto eterogeneo e difficilmente classificabile: una protezione del diritto d'autore molto evanescente<sup>123</sup>. Bisogna aggiungere, però, che le logiche che governano il rapporto tra diritto d'autore e *Big data* costringono, a parere di chi scrive, a ripensare totalmente la disciplina del diritto d'autore sia sulle banche dati - verso un approccio che non può più essere proprietario a una tipologia di dati che non sono mai né creazioni né invenzioni di coloro che li posseggono - sia nelle dinamiche di riutilizzo da parte di terzi del contenuto delle banche dati coperte dalla tutela della proprietà intellettuale.

Da questa breve disamina si evince come i *Big data* possono influenzare enormemente le dinamiche del mercato e della concorrenza e si cominciano ad intravedere i primi rilievi giuridici che coinvolgono aspetti, anche importanti, dell'architettura della disciplina a tutela della concorrenza. Sono emersi, inoltre, oltre agli aspetti descritti sopra, alcune tendenze generali di non poco momento. I *Big data* sembrano anche modificare il rapporto tra concorrenza, tutela dei consumatori e la tutela della *privacy*: la tutela della concorrenza sembra, in maniera sempre più stretta, dipendere dal rispetto delle norme a tutela del consumatore e delle norme a tutela della *privacy*. Considerando che le asimmetrie informative nascono fondamentalmente dalla capacità degli operatori del mercato di “conoscere” il più possibile i consumatori e che esse sono molto spesso alimentate da pratiche e metodologie di utilizzo dei dati che ledono alcune disposizioni di quelle norme, garantire il rispetto della *privacy* e tutelare il consumatore nel mercato diventano fondamentali per garantire il rispetto della concorrenza. Aspetti che, per alcuni studiosi, richiedono un decisivo intervento del regolatore pubblico<sup>124</sup>. Un maggiore intervento del regolatore pubblico viene lasciato trapelare anche da alcuni documenti dell'amministrazione americana<sup>125</sup>, la quale pare volere abbandonare

---

123 M. BOGNI – A. DEFANT, *Big data: diritti IP e problemi della privacy*, cit.. Anche European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.

124 Si v. l'intervento del Prof. A. ODLYZKO, *Privacy, price discrimination and seeds of capitalism's destruction*, cit.; European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.

125 Nell'Executive Office of the President of the U.S.A., *Big data: seizing opportunities, preserving values*, cit., si parla chiaramente di estendere le tutele per il consumatore garantita dal *Fair Credit Reporting Act*, che dà la possibilità al consumatore di conoscere e gestire quali i propri dati che sono in possesso delle aziende del credito, anche alle aziende di intermediazione dei dati, così come era già

politiche che permettono al mercato di autoregolarsi a favore di interventi maggiormente (anche se non eccessivamente) incisivi, soprattutto ai fini della tutela del consumatore e della *privacy*. L'amministrazione americana è certamente una delle amministrazioni che ha dimostrato di avere maggiore attenzione al tema, ma queste stesse necessità possono essere riscontrate, anche se in maniera molto meno strutturata, in alcuni documenti della Commissione europea<sup>126</sup>. La tendenza generale e desumibile da questi documenti, infatti, come vedremo anche nel paragrafo successivo, è innanzitutto quella di permettere al cittadino/consumatore di avere contezza di quanti e quali dati personali vengano utilizzati o siano in possesso di amministrazioni pubbliche e private, una tendenza che interroga il regolatore sulla necessità di integrare maggiormente, da un lato, la disciplina sulla tutela della *privacy* e quella sulla tutela del consumatore per garantire la concorrenza, dall'altro, di garantire un maggiore coordinamento tra le autorità amministrative che si occupano di concorrenza, di protezione della *privacy* dei cittadini e di telecomunicazioni.

### 3. *Big data e il profilo individuale della tutela dei dati personali.*

Oltre al rilievo giuridico che i *Big data* stanno assumendo nelle dinamiche del mercato e per la disciplina a tutela della concorrenza, è necessario sottolineare come il maggiore impatto giuridico del fenomeno dei *Big data*, come si è potuto osservare in parte quando abbiamo trattato alcuni aspetti del rapporto tra essi e la tutela del consumatore, riguarda la tutela della riservatezza e il trattamento dei dati personali. Il rapporto tra *Big data* e *privacy* è stato, di fatto, il profilo giuridico maggiormente indagato dalla letteratura scientifica fino a questo momento, non solo perché i dati personali, come abbiamo visto quando abbiamo parlato delle tipologie di utilizzo dei *Big data*, sono i dati più preziosi in quanto da essi si possono ricavare informazioni fondamentali a fini commerciali e di controllo, ma perché la *privacy* è il terreno che, in prospettiva, può essere maggiormente modificato dall'utilizzo dei *Big data*.

Il riutilizzo dei dati, anche personali, per fini differenti rispetto a quelli per cui sono stati raccolti, l'integrazione delle banche dati - contenenti informazioni personali, ma anche

---

stato suggerito dalla Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Business and Consumers*, 2012. Nello stesso documento viene anche sottolineato come sia necessario incentivare e diffondere l'utilizzo delle impostazioni *Do No Track* per tutelare il consumatore da tracciamenti continui e permettergli di limitare il tracciamento delle sue azioni on-line.

126 Si v. diffusamente COM (2014) 442, *Comunicazione della Commissione europea al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, Verso una florida economia basata sui dati*.

informazioni che non sono strettamente personali, ma riguardano aspetti della vita delle persone (si pensi ai consumi energetici, ai dati residui etc.) - e l'elaborazione dei dati contenuti al loro interno, che, di fatto, contribuiscono a rivelare informazioni personali, la tendenza a profilare in maniera sempre più completa le persone, a categorizzarle in maniera sempre più puntuale ed elaborata e a prendere delle decisioni sulla base di queste profilazioni e categorizzazioni, hanno fatto sostenere ad alcuni studiosi, che questi metodi stanno mettendo totalmente in crisi la disciplina della tutela della riservatezza<sup>127</sup>.

Il fenomeno dei *Big data* e la loro elaborazione ha fatto emergere il problema della tutela della riservatezza sotto due profili differenti: un profilo prettamente individuale e un profilo collettivo<sup>128</sup>. In questo paragrafo ci occuperemo del profilo individuale del rapporto tra *Big data* e *privacy*, cioè di quegli aspetti che attengono al rapporto tra l'individuo, i dati che gli appartengono, e le norme che questo rapporto tutelano.

La disciplina della riservatezza dei dati personali è fondamentalmente basata, sia in Europa che negli U.S.A., su dei principi generali e su strumenti di tutela - costruiti nel tempo per garantire la corretta applicazione della normativa - che non riescono più a interpretare la natura variegata dei dati, le loro nuove modalità di utilizzo e la costanza del loro reperimento e della loro elaborazione.

Prendendo come esempio la disciplina italiana sulla *privacy*, contenuta all'interno del d.lgs. n. 196/2003 (il c.d. Codice della *privacy*) - che ricalca praticamente i contenuti della direttiva n. 95/46/CE, la quale disciplina ancora la materia in tutta l'Unione europea<sup>129</sup> - le definizioni e i principi in essa contenuti, appare evidente la sua inadeguatezza.

Se la definizione di “trattamento” dei dati personali, contenuta nell'art. 4, comma 1, lett) a del d.lgs. n. 196/2003<sup>130</sup>, è ancora utile, in quanto fondamentalmente generica, per

127 Si v. M. BOGNI – A. DEFANT, *Big data: diritti IP e problemi della privacy*, cit.; A. MANTELERO, *Il futuro regolamento EU sui dati personali e la valenza “politica” del caso Google: ricordare e dimenticare nella digital economy*, in *Il Diritto dell'informazione e dell'informatica*, 2014, pag. 680; V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 206-212.

128 Nello stesso senso A. MANTELERO – G. VACIAGO, *Data protection in a Big data society. Ideas for a future regulation*, *Digital investigation*, November 2015, p. 104-109. L'a. parla, anziché di profilo individuale e profilo collettivo, di micro scenario e di macro scenario, referendosi al profilo individuale nel primo caso e al profilo collettivo nel secondo.

129 La direttiva 95/46/CE è in corso di riforma. Entro i primi mesi del 2016 verrà approvata ed entrerà in vigore il nuovo regolamento generale sulla protezione dei dati personali, il quale non sembra modificare sostanzialmente aspetti inerenti il nostro lavoro.

130 “Ai fini del presente Codice si intende per: a) qualunque operazione o complesso di operazioni, effettuati anche senza l'ausilio di strumenti elettronici, concernenti la raccolta, la registrazione, l'organizzazione, la conservazione, la consultazione, l'elaborazione, la modificazione, la selezione, l'estrazione, il raffronto, l'utilizzo, l'interconnessione, il blocco, la comunicazione, la diffusione, la cancellazione e la distruzione di dati, anche se non registrati in una banca di dati.”

comprendere le operazioni generali che si compiono con i *Big data*, nonostante nella sostanza si siano profondamente modificate nelle tecniche e nei risultati, già la definizione di dato personale non è più utile a comprendere tutte le tipologie di dati che oggi è possibile utilizzare nella loro elaborazione. Se il dato personale, secondo il medesimo articolo, è «qualunque informazione relativa a persona fisica, identificata o identificabile, anche indirettamente, mediante riferimento a qualsiasi altra informazione, ivi compreso un numero di identificazione personale» si evince che tutti i metadati e i dati residui, tipologie di dati che stanno riempiendo progressivamente ed esponenzialmente le banche dati di tutti i colossi industriali del web e non solo, rischiano di non essere compresi in questa definizione in quanto, presi singolarmente, non sono né informazioni formate e strutturate né dati. Una criticità presente nell'ordinamento europeo, ma che è percepita come tale anche nell'ordinamento statunitense<sup>131</sup>.

Anche i principi generali sul trattamento dei dati che hanno retto la disciplina sulla protezione dei dati personali fino ad ora, sembrano essere recessivi. Tornando a prendere la disciplina italiana come punto di partenza per un riferimento più ampio a livello europeo, l'art. 23 del d.lgs. n. 196/2003 prevede che nessun trattamento di dati personali possa essere posto in essere senza un previo consenso dell'interessato, il quale deve essere informato sulle modalità, sulle finalità del trattamento, sui suoi diritti e sui doveri di colui che utilizza quei dati. Il consenso, inoltre, deve essere liberamente prestato e riferibile ad uno specifico trattamento.

Il principio del consenso informato, declinato in questo modo e che appunto costituisce il caposaldo della disciplina della privacy, sia nell'ordinamento europeo che nell'ordinamento statunitense, non può più tutelare i dati personali dal fenomeno dei *Big data*. Come dicevamo in precedenza, una delle dinamiche incentivate notevolmente dai *Big data* sono proprio gli utilizzi secondari dei dati, tendenzialmente per finalità differenti rispetto a quelle per cui i dati sono stati prodotti e su cui è stato chiesto il consenso. Si capisce come il principio in esame non sia più utile a tutelare i dati personali in quanto, da un lato, gli utilizzi e le finalità secondarie non potranno mai essere tutte chiare alle imprese quando il consenso viene richiesto, e, dall'altro, non sarebbe neanche possibile richiedere il consenso per ogni singolo trattamento dei dati, come non sarebbe possibile chiedere un consenso generico sul trattamento dei dati, pena l'annullamento di qualsiasi tutela effettiva<sup>132</sup>.

---

131 Le stesse problematiche inerentemente la gestione dei metadati e il loro rapporto con la disciplina della privacy statunitense è riportata nel documento *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014.

132 In questo senso ampiamente V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 206-212.

Venuto meno il principio del consenso informato sul trattamento dei dati personali, siamo di fronte ad una vera paralizzazione della disciplina della privacy<sup>133</sup>. Una criticità che è delineata anche nei documenti sulla privacy prodotti dalla Casa Bianca negli ultimi due anni, dove vengono evidenziate le carenze di un sistema di tutela fondato sul principio del consenso informato e preventivo<sup>134</sup>. Nello stesso senso anche il Garante europeo della privacy, il quale ha recentemente sottolineato come la disciplina europea sulla protezione dei dati debba essere adeguatamente attuata, tenendo in considerazione il fenomeno in analisi<sup>135</sup>.

Anche i principi che attengono alle modalità del trattamento dei dati sembrano non riuscire più a cogliere le novità del fenomeno. L'art. 3 del d.lgs. n. 196/2003 impone un trattamento dei dati personali che rispetti il principio di necessità - il quale prevede di ridurre al minimo l'utilizzazione di questi dati, soprattutto nei casi in cui è possibile perseguire le medesime finalità anche con dati anonimi o che comunque non leghino direttamente il dato all'individuo - e in maniera tale che i dati siano raccolti e registrati per scopi determinati, espliciti e legittimi, siano esatti e aggiornati, pertinenti e completi e non eccedenti rispetto alle finalità per le quali sono state raccolte. I dati, inoltre, devono essere conservati in una forma che permette la loro associazione con una persona per un periodo strettamente indispensabile all'espletamento degli scopi per i quali sono stati raccolti.

L'elaborazione dei *Big data* e, in particolare, di quella parte dei dati che hanno un carattere personale, per sua natura non può rispettare il principio di necessità in quanto sono operazioni che per avere un'utilità ed un'efficacia devono essere costanti, in cui i dati non sono utilizzati per una sola finalità, molte volte non sono neanche dati strutturati e non sono neanche esatti. Infine la loro conservazione ha alla base una natura estremamente conservativa in quanto è intrinseco al fenomeno dei *Big data* la tendenza a conservare tutti i dati possibili.

---

133 Il termine “paralizzazione della privacy” è presente in V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 206-212. In questo senso anche A. MANTELERO, *The future of consumer data protection in the E.U. Rethinking the “notice and consent” paradigm in the new era of predictive analytics*, *Computer Law & Security Report*, November 2014, p. 643-660.

134 Queste considerazioni sono presenti sia in *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014, sia nel precedente documento *Executive Office of the President of the U.S.A., President's Council of Advisors on Science and Technology, Big Data and Privacy: a technological perspective*, Maggio 2014.

135 European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.

In ultimo è utile sottolineare come da più parti sia stata sottolineata anche l'inadeguatezza di altri strumenti classici della tutela della riservatezza come la dissociazione o l'anonimizzazione dei dati<sup>136</sup>.

Il primo strumento risulta essere molto inefficace in un contesto come quello in esame, in cui i dati provenienti dal web costituiscono la tipologia di dati più numerosa. Alcuni studiosi hanno riportato l'esempio, emblematico, di una dissociazione su Google Maps, avvenuta in Germania. Un gruppo di cittadini chiese a Google di oscurare sulle mappe di Google Maps un gruppo di appartamenti singoli con la motivazione che fossero individuabili e oggetto di possibili furti. La dissociazione di questa informazione, di un'informazione geografica, ha portato ad un risultato tutt'altro che utile per la protezione della *privacy*: l'oscuramento dell'appartamento dalle mappe di Google, di fatto, aveva reso certa ed evidente la presenza di appartamenti. Si capisce con in determinati contesti lo strumento della dissociazione non sia più utile per tutelare la riservatezza dei dati personali<sup>137</sup>.

Stesso discorso, forse in maniera ancora più pregnante, può essere svolto sull'anonimizzazione dei dati. Anonimizzare i dati significa, in un contesto come questo, depurare i *dataset* da qualsiasi dato che possa essere utilizzato per identificare una persona. Le caratteristiche dei *Big data* e i loro metodi di elaborazione, però, permettono di superare facilmente l'anonimato di un'informazione, in quanto permettono con facilità la de-anonimizzazione delle informazioni stesse. La capacità di raccogliere molti più dati rispetto al passato, e le loro tecniche di elaborazione, soprattutto quelle che combinano e fondono i dati di *dataset* differenti, permettono una facile ricostruzione di moltissime informazioni personali partendo da dati che personali non sono e una elevatissima capacità di associarle ad una persona, rendono impossibile qualsiasi tentativo di anonimizzazione<sup>138</sup>. Una criticità che è stata ampiamente segnalata

---

136 In questo senso V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 206-212; P. OHM, *Broken promises of privacy: responding to the surprising failure of anonymization*, *UCLA Law Review*, 2010, 1701; *Executive Office of the President of the U.S.A., President's Council of Advisors on Science and Technology, Big Data and Privacy: a technological perspective*, Maggio 2014.

137 L'esempio viene riportato in V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 208.

138 P. OHM, *Broken promises of privacy: responding to the surprising failure of anonymization*, *UCLA Law Review*, cit.; Interessante è l'analisi presente in PCAST Report, *Big Data and Privacy*; Harvard Law, Petrie-Flom Center, *Online Symposium on the Law, Ethics & Science of Re-identification Demonstrations*, <http://blogs.law.harvard.edu/billofhealth/2013/05/13/online-symposium-on-the-law-ethics-science-of-re-identification-demonstrations/>, in quanto si parla di “effetto mosaico” proprio per il fenomeno di incrociare dataset differenti e di ricavarne informazioni personali in maniera indiretta o attraverso deduzioni ispirate dal contenuto di queste informazioni. Esempi di de-anonimizzazioni che hanno suscitato clamore nell'opinione pubblica negli U.S.A. sono riportati in V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 208-211.

anche in documenti della Casa Bianca, in cui appunto si è sottolineato come sia proprio la natura intrinseca del fenomeno dei *Big data* e i suoi metodi di utilizzo a scardinare l'impianto della disciplina della *privacy* e, in particolare a rendere obsoleti strumenti di protezione come l'anonimizzazione<sup>139</sup>.

Un altro aspetto che sfugge alla disciplina sul trattamento dei dati personali è quella inerente il trattamento dei dati da parte di terzi, che non hanno raccolto i dati, ma li conservano o li utilizzano successivamente. È un fenomeno questo che, in realtà, si era già posto molto prima della comparsa del fenomeno dei *Big data*, ma che con esso ha conosciuto dei margini di espansione rilevanti. Il documento dell'amministrazione americana mette in evidenza, infatti come l'elaborazione a distanza dei dati, la diffusione delle tecnologie *cloud storage*<sup>140</sup>, l'utilizzo anche poco trasparente da parte di terzi (come ad esempio società di intermediazione dei dati o di amministrazioni pubbliche) dei dati che non hanno raccolto, ha messo in difficoltà la giurisprudenza che fino ad oggi aveva sistematizzato il rapporto tra *privacy* e dati affidati a terzi. Nonostante nel 2010 la United States Court of Appeals for the Sixth Circuit nel caso *United States v. Steven Warshak et al.*, infatti aveva esteso la tutela del IV emendamento anche alle comunicazioni elettroniche e al loro contenuto, la violazione da parte dei terzi della *privacy* dei cittadini è continuata, considerando i casi eclatanti di violazione da parte del governo statunitense dei dati dei cittadini e le preoccupazioni che si possono leggere in filigrana nel documento citato in merito all'eccessiva leggerezza con cui le agenzie governative utilizzano i dati sui cittadini in possesso delle aziende private e utilizzati per finalità commerciali<sup>141</sup>. Inoltre, ci si è domandato se la tutela del IV emendamento possa essere estesa ai contenuti dei metadati o a tutti i nuovi dati prodotti dalle nuove tecnologie<sup>142</sup>.

---

139 *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014.

140 Per *cloud storage* si intende un modello di conservazione di dati su computer in rete dove i dati stessi sono memorizzati su molteplici server virtuali generalmente ospitati presso strutture di terze parti o su server dedicati. Per una panoramica sul rapporto tra *privacy* e *cloud computing* si v. D. LAMETTI, *Cloud computing: verso il terzo Enclosures Movement?*, in *Riv. critica dir. Priv.*, 3, 2012, 366.

141 La vicenda a cui si fa riferimento è chiaramente quella del *Datagate*. Un programma governativo statunitense denominato PRISM aveva raccolto senza autorizzazione migliaia di informazioni personali, in particolare violandone le comunicazioni elettroniche, di cittadini statunitensi e stranieri, anche accedendo alle informazioni in possesso dei grandi provider del web. Per un commento e per le prime implicazioni giuridiche della vicenda sulla *privacy* di v. F. PIZZETTI, *Datagate, Prism, caso Snowden: il mondo tra nuova grande guerra cibernetica e controllo globale*, in [www.federalismi.it](http://www.federalismi.it), fasc. 13, 2013.; Per il resto si v. *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014. Per una disamina completa della disciplina della *privacy* nell'ordinamento statunitense si v. W. LAFAVE, *Search and Seizure: A Treatise On The Fourth Amendment*, §§ 1.1–1.2 West Publishing, 5th ed. 2011.

142 *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014.

Il rilievo giuridico e l'impatto dei *Big data* sulla disciplina della *privacy* ha reso evidente come la disciplina della tutela dei dati personali non sia più adatta allo scopo per cui essa è costruita negli ultimi vent'anni. È percepibile un forte scollamento tra la realtà - cioè la natura dei nuovi dati che possono produrre informazioni personali, la varietà dei metodi e degli utilizzi che vengono fatti dei dati personali - e la normativa - legata al singolo dato personale e al suo sporadico utilizzo. La tendenza, sia in letteratura, sia nella reportistica delle amministrazioni pubbliche, in particolare quella statunitense, è di prendere atto del fallimento dell'approccio autorizzatorio preventivo e di provare, da un lato, a spostare l'attenzione del regolatore pubblico sulla responsabilizzazione di coloro che utilizzano i dati, su dei nuovi principi che reggano le modalità di utilizzo del dato e su un controllo soltanto successivo sul rispetto della disciplina della *privacy*, dall'altro, a rendere il cittadino più informato sui dati personali in circolazione.

Da un lato, cioè, colui che utilizza i dati, una volta richiesta l'autorizzazione è libero di utilizzare, per finalità lecite, quegli stessi dati personali quante volte vuole, magari con un vincolo di coerenza con la finalità originaria o, se questo non fosse possibile, con un obbligo di miglioramento collettivo generale<sup>143</sup>. Questo sistema, ovviamente, deve essere retto da una forte capacità sanzionatoria<sup>144</sup>. Dall'altro, invece, il cittadino dovrebbe avere, in modo chiaro e trasparente, una completa panoramica di chi possiede i propri dati personali e per quali finalità vengono utilizzati. Il documento dell'amministrazione americana, infatti, non solo propone una sostanziale standardizzazione tecnologica e normativa per permettere al cittadino l'accesso controllato alle informazioni personali in possesso di soggetti pubblici o privati all'interno di determinati settori, come la sanità o l'istruzione<sup>145</sup>, ma apre alla possibilità di poter pensare ad una più ampia unificazione normativa e tecnologica, sempre ai fini di un accesso completo e controllato dei cittadini alle informazioni personali in possesso di soggetti pubblici e privati, tra la disciplina della *privacy* e quella a tutela del consumatore.

---

143 In questo senso sia V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 233-236. Nella stessa direzione anche *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014.

144 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 233-236

145 Una serie di documenti prodotti dall'amministrazione statunitense hanno nel tempo sottolineato come si sia una necessità di razionalizzare e unificare la normativa per migliorare l'accesso e il controllo sui ai propri dati. Nel settore sanitario si v. President's Council of Advisors on Science & Technology, *Realizing the Full Potential Of Health Information Technology to Improve Health Care for Americans: The Path Forward*, The White House, December, 2010. Nel settore dell'istruzione (con una particolare attenzione per la tutela della *privacy* dei minori) si v. Department of Education, *Protecting Student Privacy While Using Online Educational Services: Requirement and Best Practices*, February, 2014.

Sembra, infatti, palesarsi la tendenza ad interrogarsi, come abbiamo già detto nel paragrafo precedente, sulla necessità di integrare le due discipline, soprattutto negli aspetti che più le accomunano, cioè quelli inerenti la gestione dei dati del cittadino/consumatore e della gestione degli stessi da parte dei soggetti privati e pubblici<sup>146</sup>. Una maggiore integrazione della normativa che significherebbe anche un maggiore coordinamento tra le autorità amministrative di regolazione maggiormente coinvolte, come le autorità di regolazione della concorrenza, della tutela della *privacy* e quella delle telecomunicazioni.

#### 4. *Big data e il profilo collettivo della tutela dei dati personali.*

Il rapporto tra *Big data* e tutela della riservatezza dei dati personali ha assunto una rilevanza ulteriore, che sfugge totalmente alla disciplina della riservatezza così come la conosciamo e che trascende dalla relazione meramente individuale tra la persona e il soggetto che utilizza i dati: in cui il primo cerca di tutelare i propri dati personali cercando di limitare il loro utilizzo (anche se oramai è lo stesso cittadino, il più delle volte, a fornire alle aziende in via del tutto volontaria le informazioni personali) o, quantomeno, conoscere come questi verranno utilizzati dai terzi e il secondo cerca di utilizzarli il più possibile per migliorare i propri servizi e la propria produttività, eccellere nel mercato.

L'ulteriore profilo giuridico che il diffuso utilizzo dei *Big data* e la loro elaborazione sempre più sofisticata fa emergere, è un profilo collettivo della tutela della riservatezza<sup>147</sup>.

Le strategie di marketing per la promozione dei propri prodotti e dei propri servizi, le strategie di prezzo differenziato di cui abbiamo parlato in precedenza, la pervasività e la predittività degli strumenti di analisi dei *Big data* alla base delle decisioni stanno avendo un impatto anche sulle dinamiche sociali, in quanto questi strumenti di analisi possono contribuire a discriminare, ghettizzare e escludere individui o gruppi di persone per l'etnia di appartenenza, per il gruppo sociale di riferimento, per alcune situazioni personali.

Si pensi ad un recente lavoro compiuto analizzando i risultati delle operazioni compiute sui motori di ricerca del web, che appunto utilizzano i *Big data* come strumenti di

---

146 European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.

147 In questo senso A. MANTELERO – G. VACIAGO, *Data protection in a Big data society. Ideas for a future regulation*, cit..

selezione delle risposte alle *query* di coloro che navigano in rete. Lo studio ha evidenziato come i motori di ricerca associassero alcuni nomi di persone diffusi tra la popolazione afro-americana degli Stati Uniti, tipo Jermaine, con il termine “arresto” molte più volte di quanto lo facessero con un nome di persona diffuso tra la popolazione bianca degli Stati Uniti, tipo Geoffrey<sup>148</sup>. Un lavoro di ricerca che ha sottolineato come l'utilizzo dei *Big data* possa generare degli effetti discriminatori sociali, oltre che individuali, rilevanti, non solo perché alimentano in maniera penetrante i c.d. «filter bubbles», in cui l'eccessiva personalizzazione dei risultati delle ricerche porta l'individuo a rinchiudersi in una *enclave* ideologica e culturale, la quale viene alimentata da questi filtri, che eliminano contenuti con idee politicamente e culturalmente differenti, e che possono peggiorare l'esclusione sociale di individui o interi gruppi sociali, ma perché da queste tipologie di elaborazioni possono derivare anche decisioni che danneggiano seriamente gli individui e i gruppi sociali<sup>149</sup>. Gli esempi riportati nel capitolo precedente hanno evidenziato come le decisioni guidate dagli algoritmi e dalle inferenze statistiche possano essere capaci di accrescere le disparità economiche e sociali degli individui o semplicemente danneggiarli: gli algoritmi sono capaci di dirci quali sono le probabilità di subire un attacco di cuore, con delle possibili pesanti ricadute sui premi assicurativi sanitari o quelle di non riuscire a rimborsare un mutuo ipotecario, con la possibilità che venga negato il finanziamento<sup>150</sup>.

Gli effetti dell'uso degli algoritmi e delle tecnologie che elaborano i *Big data* per prendere delle decisioni possono essere discriminatori, anche quando l'intento è diametralmente opposto o quando non ci si aspettano risultati negativi. Si pensi al caso della città di Boston riportato in un documento prodotto dall'amministrazione statunitense<sup>151</sup>. L'amministrazione comunale della città statunitense aveva costruito un'applicazione di nome Bump, che utilizzava un accelerometro e il GPS di uno smartphone, per raccogliere dati sulle condizioni stradali, in particolare se il manto stradale fosse rovinato o ci fossero delle buche nel manto, e dava la possibilità ai cittadini di segnalarli al dipartimento lavori pubblici della città, il quale sarebbe intervenuto soltanto dopo la segnalazione da parte degli utilizzatori dell'applicazione. Coloro che lavoravano all'applicazione, però, si accorsero subito della potenziale discriminazione che alcune fasce della popolazione potevano subire: considerando che

---

148 L. SWEENEY, *Discrimination in Online Ad Delivery*, 2013, reperibile facilmente sul sito <http://dataprivacylab.org/projects/onlineads/1071-1.pdf>.

149 Executive Office of the President of the U.S.A., *Big data: seizing opportunities, preserving values*, Washington, Maggio 2014.

150 F. PASQUALE, *The Black Box Society: The Secret Algorithm Behind Money and Information*, Harvard University Press, 2014; P. DIXON – R. GELLMAN, *The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future*, *World Privacy Forum*, April 2014.

151 Executive Office of the President of the U.S.A., *Big data: seizing opportunities, preserving values*, Washington, Maggio 2014.

le fasce più povere e gli anziani non sarebbero stati in grado di utilizzare a pieno il servizio, vincolare gli interventi alle richieste dei cittadini inviate tramite gli smartphone avrebbe avuto l'effetto di indirizzare sistematicamente gli interventi verso i quartieri più ricchi della città. La soluzione fu, infatti, di allestire comunque un servizio permanente di funzionari che controllassero lo stato delle strade cittadine e che l'applicazione diventasse non il perno, ma un supporto alla loro attività.

Il profilo collettivo della tutela della riservatezza è uno degli aspetti senza dubbio più originali che l'utilizzo dei *Big data* e la loro elaborazione ha messo di fronte a coloro che hanno studiato il fenomeno. Un profilo che non può non integrarsi con le considerazioni che abbiamo fatto nel paragrafo precedente sulla necessità di una diversa concezione della disciplina della *privacy* nell'era dei *Big data*. Un approccio che deve ribaltare totalmente l'impostazione dell'attuale disciplina della *privacy* e che obbliga a ripensarla anche con l'obiettivo di ricomprendere tutele anche per dinamiche collettive di lesione della tutela della *privacy*, che qualche autore ha già sottolineato<sup>152</sup>.

La categorizzazione di massa, la profilazione penetrante, le analisi probabilistiche e predittive di fenomeni sociali o comportamenti individuali, oltre a fare sorgere un nuovo profilo giuridico, in chiave collettiva, della tutela della riservatezza, possono fare sorgere anche interrogativi molto più generali e che coinvolgono direttamente garanzie tutelate dalle Carte costituzionali.

L'elaborazione dei *Big data*, come l'abbiamo descritta in questo paragrafo, può ledere profondamente anche il principio di eguaglianza. Un ragionamento per cui la possibilità che una persona possa compiere determinate scelte o assumere determinati comportamenti semplicemente perché possiede alcune caratteristiche o perché appartiene a determinati gruppi sociali o a determinate etnie o semplicemente perché così ha previsto il risultato di un'analisi algoritmica con i *Big data*, può produrre delle forti discriminazioni per associazione che penalizzano la persona, non solo perché valutata al netto della propria volontà personale, ma soprattutto in quanto facente parte di un gruppo sociale o di una determinata etnia<sup>153</sup>.

Una prospettiva che per qualche studioso, anche portando la questione ad un livello più filosofico, disumanizza le scelte individuali, le collettivizza, cancella il libero arbitrio e rischia di dare valore causale ad analisi ed elaborazioni che questo valore, come

---

152 Si v. A. MANTELERO – G. VACIAGO, *Data protection in a Big data society. Ideas for a future regulation*, cit.; In parte anche European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.

153 Per un riferimento sulle discriminazioni per associazione dovute alle analisi predittive si v. B.E. HARCOURT, *Against Prediction: Profiling, policing and punishing in an actuarial age*, University of Chicago Press, Chicago, 2006.

abbiamo avuto modo di vedere, non hanno in alcun caso, considerando che si tratta di correlazioni<sup>154</sup>.

La collettivizzazione delle azioni individuali e l'appiattimento delle possibili scelte personali sulle probabilità che queste possano accadere, considerando determinate caratteristiche o determinati avvenimenti che riguardano la persona o il gruppo sociale a cui essa appartiene, lede anche il principio personalistico che, alcune Carte costituzionali, come la nostra, considerano una dei capisaldi di tutto l'impianto costituzionale e contenuto nell'art. 2 della Costituzione. Se, infatti, il singolo è visto come una risorsa per la collettività e lo sviluppo della persona è un obiettivo costituzionale, che ritroviamo anche in altre disposizioni della Costituzione italiana (nell'art. 3 precisamente), logiche che vedono i dati come qualcosa di distaccato dalla persona a cui questi dati appartengono o che tendono, invece, a massificare l'individuo e ad annichilire la libertà di scegliere i propri comportamenti e la responsabilità delle proprie azioni, vanno nella direzione opposta<sup>155</sup>.

Utile per la ricostruzione di un nuovo approccio alla privacy e alla sua disciplina, che abbia come obiettivo fondamentale quello di riconnettere il profilo individuale con il profilo collettivo della tutela della riservatezza e che abbia tra i suoi elementi una maggiore consapevolezza da parte del cittadino/consumatore sulle sue informazioni in circolazione è anche dare una certa rilevanza ad un altro aspetto di un certo rilievo giuridico che i *Big data* stanno esasperando.

Oltre alla tendenza a disumanizzare le azioni individuali, c'è una chiara tendenza a disumanizzare i dati e questo è evidente dalle pratiche di utilizzo dei dati personali che abbiamo illustrato in questi paragrafi. I dati - in particolare quelli personali, intesi qui nel senso più ampio del termine e cioè riguardanti il corpo in sé o riguardanti più in generale il comportamento dei cittadini - non vengono considerati come delle astrazioni personali, ma come elementi asettici da elaborare per garantire meglio il fine che ci si è posti con l'utilizzo dei *Big data* (una maggiore produttività dell'azienda, nei casi riportati in questo paragrafo). Il tema del corpo elettronico, cioè l'insieme delle informazioni

---

154 In questo senso si v. V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 219.

155 Il riferimento di partenza è la dottrina che progressivamente ha messo in primo piano il principio personalistico e l'individuo come chiave di una lettura costituzionalmente orientata del rapporto tra cittadino e amministrazione. Si v. U. ALLEGRETTI, *Amministrazione pubblica e Costituzione*, Padova, 1991; V. BACHELET, *Costituzione e amministrazione. Scritti giuridici*, Roma, 1992; G. PASTORI, *Diritti e servizi oltre la crisi dello Stato sociale*, in *Scritti Ottaviano*, Milano, 1993; F. BENVENUTI, *Il nuovo cittadino. Tra libertà garantita e libertà attiva*, Venezia, 1994; G. ARENA, *L'utente-sovrano*, in *Studi in onore di Feliciano Benvenuti*, Modena, 1996; S. BARTOLE, *Interpretazioni e trasformazioni della Costituzione repubblicana*, Bologna, 2004; R. CAVALLO-PERIN, *La configurazione della cittadinanza amministrativa*, in *Dir. amm.*, 2004, 202 ss.

riguardanti il corpo e il comportamento di una persona<sup>156</sup>, il “corpo” di informazioni che ci riguardano personalmente e che sono sparse principalmente nella rete e che sono a disposizione di terzi, costituisce un'altra rilevante implicazione giuridica dei *Big data*. Con il fenomeno dei *Big data* il corpo elettronico di una persona sembra definirsi chiaramente, in maniera ancora più dettagliata, ma che non identifica sempre in maniera corretta la persona nel mondo digitale. La pervasività della datizzazione delle persone e dei comportamenti, la pervasività dell'elaborazione dei *Big data* ha portato in generale ad una crescita esponenziale della quantità di informazioni che riguardano una persona, che sono nella disponibilità di vari soggetti<sup>157</sup>. Le nuove tecnologie avevano già da tempo influenzato i modi in cui l'identità individuale veniva costruita ed articolata, in particolare nel mondo digitale. L'identità individuale può essere scomposta e ricostruita a seconda dei vari contesti in cui si opera, a seconda delle varie tipologie di dati che si utilizzano e non è un dato preesistente, ma è il prodotto di un processo continuo di costruzione<sup>158</sup>. Questo carattere dinamico dell'identità è stato però esasperato dai *Big data*, oltre ogni limite, al punto tale che la raccolta delle informazioni è ormai in grado di produrre effetti addirittura senza bisogno di ulteriori interventi<sup>159</sup>.

Le questioni legate al corpo digitale e all'identità personale dentro il mondo digitale sono decisive, soprattutto dal punto di vista culturale, per affrontare le nuove sfide, in particolare di natura giuridica, che i *Big data* pongono in riferimento alle tematiche riguardante la privacy dei cittadini e della sua tutela.

##### 5. *Big data, sorveglianza e prevenzione ai fini di sicurezza e controllo.*

L'utilizzo dei *Big data* per profilare e categorizzare gli individui e i gruppi sociali e, sulla base di queste profilazioni e categorizzazioni, prendere delle decisioni non ha

---

156 Questa definizione di corpo digitale è presente in S. RODOTÀ, *Il diritto di avere diritti*, Bari, 2012, p. 26 ss.

157 In questa prospettiva sono interessanti le considerazioni di D. LYON, *Massima Sicurezza. Sorveglianza e “guerra al terrorismo”*, trad. it. E. Greblo, Milano, 2005, pp. 163.164. Sul tema del corpo digitale la dottrina giuridica si è espressa ampiamente. Per tutti si v. S. RODOTÀ, *Of Machine and men*, in M. HILDEBRANDT, A. ROUVROY (A CURA DI), *Law, Human Agency and Autonomic Computing: The Philosophy of Law Meets the Philosophy of Technology*, New York, 2011, 179 ss.; G. MARINI, *Diritto alla privacy. Commento al Codice in materia di protezione dei dati personali*, in A. BARBA – PAGLIANTINI (A CURA DI), *Delle persone*, in E. GABRIELLI (DIRETTO DA), *Commentario del codice*, UTET, 2013, p. 199 e ss..

158 In questo senso già G. RESTA, *Identità personale e identità digitale*, in *Diritto dell'informazione e dell'informatica*, 2007, p. 511.

159 Non in riferimento al fenomeno dei *Big data*, ma in generale all'avvento di nuove tecnologie che permettono di elaborare meglio i dati si v. G. MARINI, *Diritto alla privacy. Commento al Codice in materia di protezione dei dati personali*, cit.

coinvolto soltanto le aziende private per finalità commerciali, ma ha coinvolto anche il settore pubblico.

L'elaborazione dei *Big data* ha fatto emergere aspetti giuridici rilevanti all'interno delle questioni inerenti il controllo e la prevenzione. Dagli esempi che sono stati fatti nel capitolo precedente, l'elaborazione dei *Big data* viene utilizzata sempre più spesso per controllare in maniera capillare il territorio, le persone o gruppi sociali, per prevenire il terrorismo o, in via generale, la commissione di reati e crimini. Molti programmi governativi statunitensi utilizzano i *Big data* per riconoscere potenziali terroristi (come i programmi dell'NSA e il programma del Dipartimento della Homeland Security FAST di cui abbiamo parlato nel capitolo precedente) e per prevedere quali sono i soggetti o i luoghi dove è probabile che si verifichino eventi criminosi (come il programma BLUE CRASH della polizia di Richmond).

Il fenomeno della sorveglianza preventiva<sup>160</sup> dello Stato sui propri cittadini per prevenire eventi criminosi, in particolare attacchi terroristici, non nasce con i *Big data*, ma è un fenomeno che, negli ultimi anni, ha modificato parzialmente i propri caratteri, accentuando il carattere invasivo e penetrante sui cittadini, proprio grazie all'utilizzo e all'elaborazione dei *Big data*. Alcuni studiosi hanno sottolineato come la sorveglianza non sia più un fenomeno autoritario (che impone comportamenti o regole), centralizzato (costruito attorno all'apparato statale in funzione di controllo e di prevenzione) e fondamentalmente intrusivo (che mira a conoscere tutto sui cittadini), ma sia diventato un fenomeno diffuso (alimentato sia dagli Stati, in funzione di controllo e di prevenzione, sia dalle aziende private, per finalità prevalentemente commerciali o legate al perseguimento di una migliore efficienza), volontaristico (in quanto i cittadini non vengono costretti a fornire delle informazioni, ma il più delle volte le forniscono volontariamente, principalmente attraverso la rete), intrusivo, ma, in maniera del tutto innovativa, anche discriminatorio (in quanto oltre a ledere la privacy dei cittadini, ha forti tendenze a categorizzarli, profilarli e, quindi, discriminarli in quanto possessori di determinate caratteristiche psico-fisiche o in quanto appartenenti a determinati gruppi sociali)<sup>161</sup>. Oggi la sorveglianza è diventata una sorveglianza “liquida”<sup>162</sup>, la quale utilizza degli strumenti e delle metodologie di elaborazione che hanno visto esplodere la loro efficacia grazie ai *Big data*: si pensi ad esempio agli impianti di sicurezza e di

---

160 Con il termine sorveglianza ci si riferisce alle modalità consuete con cui certe organizzazioni (in questo caso, pubbliche), interessate a influenzare, gestire o controllare determinate persone o gruppi, focalizzano l'attenzione sui dettagli personali. Per questa definizione si v. D. LYON, *Massima Sicurezza. Sorveglianza e “guerra al terrorismo”*, cit., p. XVI e pp. 3-5.

161 Una ricostruzione di questo genere è possibile maturarla leggendo sia D. LYON, *Massima Sicurezza. Sorveglianza e “guerra al terrorismo”*, cit., p. XVI e pp. 3-5, che Z. BAUMAN, D. LYON, *Sesto potere. La sorveglianza nella modernità liquida*, trad. it. M. Cupellaro, Bari, 2014.

162 Si parla di sorveglianza liquida in Z. BAUMAN, D. LYON, *Sesto potere. La sorveglianza nella modernità liquida*, trad. it. M. Cupellaro, Bari, 2014.

videosorveglianza o ai rilevatori biometrici, i quali - grazie alle nuove capacità tecnologiche – sono capaci di produrre molti più dati (e i “signori dei dati” sono capaci di archiviarne in quantità sempre maggiore); ai *social network*, attraverso cui è possibile reperire informazioni e fare analisi dei network sui cittadini, prevalentemente con il loro consenso, ma in alcuni casi anche creando delle forti perplessità sulla legalità delle loro utilizzazioni<sup>163</sup>; alle metodologie che fondono i *dataset* più disparati per categorizzare, profilare collettivamente e individualmente, perfino per individuare preventivamente presunti terroristi o criminali.

Tali strumenti, che da tempo sono utilizzati dagli Stati per queste finalità e che abbiamo in precedenza classificato come fonti di produzione di *Big data* o descritto come tecniche principali di elaborazione dei *Big data*, stanno incrementando notevolmente il potere informativo dello Stato. Un potere informativo, inoltre, che è cresciuto fortemente anche grazie alla possibilità di integrare le informazioni e gli strumenti di elaborazione propri con le informazioni delle aziende private, come quelle del web, che in più di un'occasione si sono dimostrate disponibili a fornire agli Stati informazioni sui cittadini<sup>164</sup>. Un potere informativo che rischia di mettere il cittadino in forte soggezione rispetto alle istituzioni pubbliche, a causa del profondo divario informativo tra i due soggetti. Non bisogna pensare, però, che questo divario tra lo Stato e i cittadini venga prodotto soltanto dall'utilizzo dei *Big data* per la sorveglianza preventiva, ma in via principale dalla possibilità da parte dello Stato di possedere e produrre delle informazioni che i cittadini non hanno, dalla possibilità per lo Stato di diffondere determinate informazioni anziché altre.

---

163 Si pensi alla vicenda degli esperimenti sociali sul possibile cambiamento di umore degli utenti di Facebook in base alle notizie positive o negative condivise dagli amici venuta alla ribalta dell'opinione pubblica nell'estate del 2014. L'esperimento, in realtà, risale al gennaio del 2012 e ha esposto 689mila utenti a contenuti per la maggior parte positivi, in un primo caso, e soprattutto negativi, nel secondo. I risultati hanno mostrato come, innanzitutto, limitare la condivisione di emozioni sia in un senso sia nell'altro porti a una riduzione ulteriore di contenuti di questo tipo e che sentimenti positivi ci portano ad esprimerne di altrettanto ottimisti mentre vedere amici e conoscenti di cattivo umore ci condiziona in questo senso. La vicenda suscitò numerose polemiche perché l'esperimento era stato condotto, secondo i critici, senza rendere edotti gli utenti soggetti a questi esperimenti. Per ulteriori informazioni [http://www.corriere.it/tecnologia/social/14\\_giugno\\_30/facebook-l-esperimento-utenti-dispiaciuti-l-ansia-causata-7e7ff33e-003f-11e4-9185-2e4a12f9e1bf.shtml](http://www.corriere.it/tecnologia/social/14_giugno_30/facebook-l-esperimento-utenti-dispiaciuti-l-ansia-causata-7e7ff33e-003f-11e4-9185-2e4a12f9e1bf.shtml)

164 Si pensi allo scandalo del Datagate in cui i provider del web fornirono l'accesso ai propri dati alle agenzie dell'amministrazione statunitense per prevenire il terrorismo. Sul Datagate e le sue implicazioni giuridiche si v. F. PIZZETTI, *Datagate, Prism, caso Snowden: il mondo tra nuova grande guerra cibernetica e controllo globale*, cit.. La costante collaborazione tra i colossi del web e i governi non nasce con i Big data. Ne parlano ampiamente, V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 246-249; E. MOROZOV, *L'ingenuità della rete. Il lato oscuro della libertà di internet*, Torino, 2011, pp. 205-206; D. LYON, *Massima Sicurezza. Sorveglianza e “guerra al terrorismo”*, cit., p. 95 ss.

Il potere informativo dello Stato in funzione di controllo e di prevenzione, rischia di ledere profondamente i diritti fondamentali di questi ultimi.

L'utilizzo dei *Big data* per prevedere la possibilità che una persona possa compiere determinate scelte o assumere determinati comportamenti semplicemente perché possiede alcune caratteristiche o perché appartiene a determinati gruppi sociali o a determinate etnie o semplicemente perché così ha previsto il risultato di un'analisi algoritmica con i *Big data*, può produrre delle forti discriminazioni per associazione che ledono profondamente il principio di eguaglianza. Un aspetto che senza dubbio avvicina le dinamiche sottostanti la sorveglianza preventiva a quello che abbiamo chiamato profilo collettivo della tutela della riservatezza: certamente le finalità sono molto differenti, ma gli effetti sui cittadini possono essere simili.

Molti studiosi hanno evidenziato come la sorveglianza preventiva frutto dell'elaborazione dei *Big data*, oltre a ledere il principio di eguaglianza, potrebbero ridurre fortemente le libertà civili dei cittadini, tutelate all'interno delle Carte costituzionali<sup>165</sup>. La cultura del controllo, del sospetto e della segretezza in merito all'utilizzo dei dati e delle informazioni dei cittadini, personali o meno, il potenziamento degli strumenti di controllo attraverso l'utilizzo dei *Big data*, ha prodotto sia una generalizzata lesione della segretezza dei dati e delle informazioni dei cittadini, sia la violazione di libertà costituzionali, come quella sull'inviolabilità delle comunicazioni.

Anche l'amministrazione statunitense, dopo lo scandalo del *Datagate*, ha cominciato ad affrontare queste criticità. Nel documento che tenta di avere un primo approccio complessivo al tema dei *Big data*, infatti, l'amministrazione sottolinea in più parti come l'utilizzo dei *Big data* per questioni legate alla sicurezza possano essere lesive delle libertà garantite costituzionalmente ai cittadini. Il documento sottolinea, poi, come l'amministrazione statunitense abbia deciso di affrontare queste criticità e tutelare maggiormente i diritti dei cittadini. Dal 2012, infatti, l'utilizzo e l'approvazione definitiva di nuovi programmi di sorveglianza che vengono costruiti dal Dipartimento della Homeland Security è vincolata a delle analisi di impatto sulla *privacy* e le libertà dei cittadini sui programmi di sorveglianza<sup>166</sup>. I nuovi protocolli contengono due strumenti che, secondo l'amministrazione statunitense, dovrebbero rendere più chiare e trasparenti le procedure di utilizzo dei dati, anche personali, dei cittadini statunitensi (ma anche dei cittadini stranieri) e garantire un controllo generale su chi utilizza i dati e per quali finalità. Il primo strumento è una nuova procedura di autorizzazioni multilivello che permette di accedere e utilizzare determinate informazioni a

---

165 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 237 ss.

166 *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014.

determinate persone: più sono informazioni delicate, meno persone sono autorizzate ad utilizzarle. Il secondo strumento è un sistema di *tags* sui dati che monitora il percorso che determinati dati, meritevoli di tutela, fanno quando vengono utilizzati durante le operazioni, di monitorare chi e perché utilizza quei dati<sup>167</sup>.

La cultura della sorveglianza preventiva rischia, inoltre, di ledere pesantemente il principio di presunzione di innocenza, un principio cardine dei nostri sistemi giudiziari. La volontà di molte amministrazioni di utilizzare programmi che elaborano i *Big data* per prevenire la commissione di attacchi terroristici o di semplici atti criminali oppure la volontà di comprendere il profilo di coloro potrebbero commettere determinati reati o, ancora, la capacità di fare delle analisi predittive su quante probabilità ci sono che un determinato individuo commetta un reato, i quali non sono più considerabili come scenari futuristici visti gli esempi che abbiamo riportato nel paragrafo precedente, potrebbero portare le autorità di sicurezza a limitare la libertà di comunicazione o la libertà di movimento o, nel peggiore dei casi, la libertà *tout court*, di determinati cittadini solo sulla base dell'elaborazione dei *Big data*.

Molti studiosi hanno sottolineato anche come questo modo di approcciarsi alla responsabilità penale dei cittadini si allontana anche dai principi costituzionali in merito alla funzione rieducatrice della reclusione di coloro che hanno commesso dei reati<sup>168</sup>. Il welfarismo penale - che indicava la commissione del crimine come sintomo di deprivazione sociale o indice di povertà e la privazione della libertà personale come momento di rieducazione, al fine di reintegrare il cittadino nella società, che tradizionalmente ha caratterizzato l'approccio alla detenzione nelle democrazie occidentali - lascia il passo ad un approccio alla reclusione e alla commissione dei reati tutto incentrato sulla asettica e amorale osservazione della situazione di fatto o degli assetti sociali con il fine di minimizzare, attraverso l'utilizzo della tecnologia, il rischio di commettere reati che possano ledere l'equilibrio sociale<sup>169</sup>. Questa nuova concezione della responsabilità penale, che i sociologi chiamano teoria “tardomoderna” del controllo sociale, sembra del tutto incompatibile con le migliori tradizioni costituzionali sul punto, come la nostra, che hanno sposato chiaramente il welfarismo penale<sup>170</sup>.

---

167 Department of Homeland Security, *Privacy Impact Assessment for the Neptune Pilot*, September 2013; Department of Homeland Security, *Privacy Impact Assessment for the Cerberus Pilot*, November, 2013.

168 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 213 ss.

169 Per una disamina delle categorie utilizzate qui per descrivere le tipologie di controllo del crimine si v. D. GARLAND, *The Culture of Control*, University of Chicago Press, Chicago, 2001, pp. 182-190.

170 Per approfondire, dal punto di vista sociologico, le teorie sul controllo del crimine e per le prime considerazioni sugli effetti dell'utilizzo di grandi volumi di dati si v. D. LYON, *Massima Sicurezza. Sorveglianza e “guerra al terrorismo”*, cit., pp. 110 - 115.

## 6. *Big data, potere e asimmetrie informative.*

L'utilizzo e l'elaborazione dei *Big data* per la sorveglianza preventiva è un esempio settoriale di come possa essere rilevante il potere informativo dello Stato, in questo momento storico. La possibilità di reperire e produrre dati sulla realtà, la capacità di poter accedere anche a dati e informazioni che non sono direttamente in suo possesso e utilizzarli per governare, può produrre effetti rilevanti sui cittadini, come abbiamo potuto vedere nel caso delle funzioni di controllo e prevenzione per garantire la sicurezza pubblica. Il rilievo giuridico dei *Big data* nel settore pubblico, infatti, è fortemente legato con il potere informativo dello Stato e con gli effetti che esso produce, in primo luogo, sui cittadini, ma anche su tutto il sistema democratico.

Fino a questo momento, sia quando abbiamo parlato di sorveglianza preventiva, sia quando abbiamo parlato di potere informativo, abbiamo fatto riferimento allo Stato in senso più generale. In realtà le attività conoscitive e il potere informativo, che dalle prime deriva, è storicamente un'attività in capo al potere esecutivo: le amministrazioni pubbliche raccolgono i dati, producono le informazioni per prendere decisioni e per permettere agli organi politici di prendere delle decisioni. Questa considerazione non può non mettere in rilievo come la capacità di datizzare la realtà e di trasformare questi dati in informazioni, che i *Big data* mettono a disposizione del potere esecutivo, possa incrinare gli equilibri democratici tra poteri. In un contesto di questo genere il Parlamento, che legifera e controlla il governo sulla base di informazioni che, in gran parte, provengono da fonti del governo stesso e delle sue amministrazioni, si può trovare in una posizione fortemente subordinata al potere esecutivo. Questa prospettiva sugli effetti dei *Big data* sul settore pubblico non è ancora stata indagata dalla dottrina, ma è possibile trovare dei riferimenti utili che sottolineano quali sono le difficoltà informative che il Parlamento si trova a dover affrontare. Prenderemo come riferimento esemplificativo il Parlamento italiano, ma è le considerazioni che si faranno possono essere certamente estese a tutte gli ordinamenti liberal-democratici come il nostro.

Qualche studioso, infatti, ha messo in rilievo come il legislatore, si trovi in maniera sempre crescente di fronte ad una realtà che, per essere governata, ha bisogno di decisori con delle grosse competenze tecniche e capacità decisionali rapide<sup>171</sup>. Questa

---

<sup>171</sup> In questo senso e in una prospettiva nazionale si v. F. PACINI, *Dall'autorità della legge all'autorevolezza delle Assemblee*, in *Quad. cost.*, fasc. 1, 2015, pp. 9 e ss., ma anche P. CARETTI – M. MORISI, *Parlamento e politiche pubbliche*, in *Quad. cost.*, fasc. 3, 2001, pp.491 ss. Per un riferimento più internazionale, in particolare per uno spaccato sulla questione negli U.S.A. si v. diffusamente C. R. SUNSTEIN, *Simpler: the future of government*, New York, Simon & Schuster, 2013.

situazione - insieme alle criticità e alle patologie della legge e, per suo tramite, del rapporto tra esecutivo e legislativo<sup>172</sup> - ha, nel tempo, contribuito enormemente ad indebolire il ruolo decisionale del Parlamento. Ci si è trovati di fronte alla necessità, sempre più incalzante, di un'istruttoria, precedente alla decisione, che fosse curata da organi specializzati e direttamente operativi, la quale ha favorito la formazione dello squilibrio a favore delle maggiori capacità conoscitive e tecniche dei Governi<sup>173</sup>. A questa storica criticità bisogna aggiungere le potenzialità dei *Big data* - in particolare quelle in merito alla capacità di conoscere ancora meglio la realtà che il potere esecutivo è chiamato a governare e alle capacità sottese all'elaborazione dei *Big data*, come quelle analitico-descrittive e quelle predittive – le quali sarebbero totalmente nelle mani dell'esecutivo e delle amministrazioni pubbliche.

Il *gap* informativo storico tra governo e Parlamento, che rischia di approfondirsi ulteriormente con le capacità di archiviazione e di elaborazione dei *Big data*, costituisce certamente un punto di indubbio rilievo giuridico, che attiene direttamente agli equilibri tra poteri negli ordinamenti democratici. Qualche studioso, però, anche se non in termini così diretti, ha già sottolineato come i *Big data* e la loro elaborazione, possono costituire anche la soluzione ad un disequilibrio informativo così rilevante, sottolineando come il recupero della centralità del Parlamento, elemento cardine delle democrazie rappresentative, si possa ritrovare anche ripensando l'attività conoscitiva del Parlamento in un'ottica di piena autonomia conoscitiva dagli esecutivi e dalle amministrazioni<sup>174</sup>.

Il potere informativo dello Stato, come abbiamo potuto constatare anche nei paragrafi precedenti, non è l'unico potere informativo che viene favorito dai *Big data* e, forse, non è neanche il più forte potere informativo che è possibile rilevare al giorno d'oggi.

Il primo grande rilievo giuridico che si pone a chi si confronta con il tema dei *Big data*, alla base di tutte le tematiche che abbiamo trattato in questo capitolo, è certamente legato agli effetti che il loro utilizzo produce nei rapporti che intercorrono tra coloro che possiedono o possono reperire facilmente enormi quantità di dati e, di conseguenza, di

---

172 Per una valutazione sulle patologie della legge, intesa come «precipitato di una valutazione parlamentare o frutto di una collaborazione con la collegata autorità esecutiva» (definizione tratta da F. PACINI, *Dall'autorità della legge all'autorevolezza delle Assemblee*, cit., p. 11) si v. A. PIZZORUSSO, *La manutenzione del libro delle leggi ed altri studi sulla legislazione*, Torino, 1999; M. DOGLIANI (A CURA DI), *Il libro delle leggi strappato e la sua manutenzione*, Torino, 2012; B. G. MATTARELLA, *La trappola delle leggi. Molte, oscure e complicate*, Bologna, 2011; R. ZACCARIA, *Fuga dalla legge? Seminari sulla qualità della legislazione*, Brescia, 2011.

173 Oltre a F. PACINI, *Dall'autorità della legge all'autorevolezza delle Assemblee*, cit., p. 16 si v. anche U. De Siervo, *Ma chi fa i decreti legislativi?*, in V. COCOZZA – S. STAIANO (A CURA DI), *I rapporti tra Parlamento e Governo attraverso le fonti del diritto*, I, Torino, 2001 p. 497. In una prospettiva più politologica si v. S. VASSALLO, *Liberiamo la politica. Prima che sia troppo tardi*, Bologna, 2014.

174 F. PACINI, *Dall'autorità della legge all'autorevolezza delle Assemblee*, cit., pp. 25-26

informazioni e coloro che possiedono pochi dati o che non possono reperirli facilmente e, di conseguenza, possiedono meno informazioni.

La complessità del fenomeno dei *Big data* contribuisce notevolmente all'apparizione di pochi soggetti pubblici e privati che hanno il privilegio di possedere ed elaborare grandi quantità di dati<sup>175</sup>. Lo Stato, i provider del web (Google, Facebook etc.), le società di intermediazione che raccolgono informazioni per conto di terzi (società di intermediazione creditizia o quelle che svolgono analisi di mercato) sono avvantaggiati da due grandi fattori<sup>176</sup>. Il primo è quello di gestire, grazie alle loro attività economiche o istituzionali, enormi flussi informativi che vanno ad alimentare il loro patrimonio informativo. Questi soggetti, infatti, possono utilizzare i dati in via del tutto esclusiva, in quanto, il più delle volte, sono proprietari dei dati e, conseguentemente, possono accedere soltanto loro alle basi di dati in cui essi sono conservati. Il secondo fattore è quello tecnologico. Non è raro, infatti, che questi soggetti, grazie all'accesso alle nuove tecnologie, siano gli unici che abbiano la possibilità non solo di archiviare i *Big data*, ma principalmente di elaborarli e ricavarne nuove e costanti informazioni, sfruttando al meglio le loro potenzialità conoscitive.

Il vantaggio informativo di questi pochi soggetti, che qualche studioso ha soprannominato i “signori dei dati”<sup>177</sup>, crea inevitabilmente un potere informativo rilevante in capo a questi soggetti e, conseguentemente, delle forti asimmetrie informative.

Un primo tipo di asimmetria informativa è, innanzitutto, quella che si forma tra coloro che posseggono ed elaborano i dati e le informazioni e coloro che questi dati e queste informazioni ne posseggono in quantità limitata. Abbiamo visto nel corso del paragrafo come queste asimmetrie - tra le aziende private, tra le aziende private ed gli utenti/consumatori o tra lo Stato e i cittadini - possano avere delle implicazioni giuridiche rilevanti, sulla concorrenza, sulla disciplina a tutela del consumatore e sulla disciplina della privacy, ma anche sulle libertà e sui diritti dei cittadini.

Un secondo tipo di asimmetrie informative, invece, sono quelle che intercorrono tra coloro che possiedono le informazioni, tra i “signori dei dati”, cioè tra gli Stati o tra gli Stati e i provider del web.

---

175 Chiaro sulla ricostruzione degli assetti di potere e delle asimmetrie informative che si formano per effetto dell'utilizzo dei *Big data* MANTELERO A., *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, cit..

176 Esplica bene quali sono i fattori che permettono ai maggiori possessori dei *Big data* di avere un così forte potere informativo diffusamente MANTELERO A., *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, cit..

177 MANTELERO A., *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, cit., p. 135.

Se il primo tipo di asimmetrie sono state già indagate nel corso del capitolo, il secondo tipo di asimmetrie merita di essere approfondito, anche perché ha suscitato un certo interesse tra gli studiosi che se ne sono occupati<sup>178</sup>.

Il rilievo giuridico più evidente, se si indagano i rapporti e le dinamiche che si formano all'interno delle asimmetrie informative tra gli Stati e tra gli Stati e i provider del web, attiene alle reciproche influenze e agli effetti che queste influenze possono avere sulla sovranità statale.

Questo tipo di asimmetrie, in molti casi, rafforzano il potere informativo dello Stato mettendolo in una condizione di superiorità informativa sia nei confronti degli altri Stati, sia nei confronti delle grandi aziende del *web*, mentre altre volte, invece, debilitano la sovranità dello Stato stesso, che, in determinate situazioni, si trova ad essere subordinato ad altri Stati, in possesso di un potere informativo superiore, o “catturato” dai provider del *web*.

Una caratteristica peculiare del potere informativo nell'era dei *Big data* è che esso è tendenzialmente geolocalizzato, sotto il duplice profilo della nazionalità di coloro che lo detengono e del luogo fisico dove sono concentrate le strutture che raccolgono e gestiscono questi dati<sup>179</sup>. La geolocalizzazione dei dati può influire pesantemente sui rapporti tra gli Stati: si pensi alle differenti leggi applicabili sulla loro gestione e sulle tutele dei cittadini; agli effetti sui reciproci rapporti prodotti dalla concentrazione dei dati e dalla capacità di utilizzare determinate tecnologie (grandi *data center*, strumenti di analisi dei dati sofisticati, sistemi di *cloud computing* etc.) in capo ad uno Stato e all'influenza che questo può avere sullo Stato in possesso di meno dati e di meno tecnologia<sup>180</sup>.

Un esempio concreto di disequilibrio informativo è quello tra gli Stati Uniti e l'Unione europea, considerando che i maggiori provider del web sono statunitensi e che la gran parte dei dati sono concentrati negli U.S.A.. Non sorprende, infatti, come proprio in merito alla gestione dei dati l'Unione europea abbia deciso di modificare profondamente la disciplina sulla protezione dei dati europei, estendendone la portata e, quindi, la tutela in qualsiasi posto essi si trovino e non sorprendono nemmeno le vicende politiche che hanno visto contrapporsi l'Unione europea e gli Stati Uniti proprio su questo punto.

---

178 Per tutti si v. diffusamente A. MANTELERO, *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, cit.; V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 246-249;

179 A. MANTELERO, *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, cit., p. 139.

180 Ibidem.

Oppure si pensi alla tendenza che alcune aziende del web hanno a monopolizzare il settore e a influenzare o “catturare” il decisore pubblico in merito alle scelte che questo può compiere in merito alla regolazione dei mercati o a l'utilizzo dei *Big data* in generale<sup>181</sup>. Anche qui vale la pena ricordare la pressione politica che Google provò a fare sui regolatori europei dopo la sentenza della Corte di giustizia dell'Unione europea sul caso Google Spain. Dopo la sentenza, Google decise di adempiere alla pronuncia della Corte con misure di forte impatto mediatico con lo scopo di dimostrare come la sentenza della Corte fosse eccessivamente onerosa per i gestori dei motori di ricerca e che ledesse il diritto all'informazione dei cittadini. L'azienda della *Silicon Valley* ha creato uno spazio *on-line* per consentire a tutti gli utenti di inoltrare una richiesta di cancellazione delle informazioni (in applicazione del diritto all'oblio) e soprattutto rispose positivamente alle richieste di coloro che chiedevano di eliminare dai risultati della ricerca articoli delle più famose testate giornalistiche anglosassoni. L'obiettivo non dichiarato era quello di utilizzare l'opinione pubblica per fare pressione, in senso ovviamente opposto rispetto alla sentenza della Corte, sul regolatore europeo, che aveva deciso di sposare nei nuovi regolamenti, che avrebbero disciplinato la protezione dei dati personali, proprio gli orientamenti della Corte di giustizia<sup>182</sup>.

Due esempi che dimostrano come il disequilibrio informativo tra chi possiede più o meno dati e informazioni, possa ridimensionare o influenzare notevolmente la sovranità, in particolare quella statale, intesa come capacità di poter fare determinate scelte senza essere influenzati da interessi privati o interessi esteri.

Lo Stato, però, nei rapporti con i “signori dei dati” privati, non sempre si trova in una posizione totalmente subordinata. La cattura del decisore pubblico viene bilanciata dalla disponibilità da parte delle aziende del *web* a concedere senza troppi scrupoli etici o morali i dati in loro possesso, anche quelli personali dei cittadini. Alcuni studiosi, infatti, hanno riportato esempi in cui le aziende del web, non solo hanno favorito i governi democratici intenti ad utilizzare i dati in funzione di controllo e di prevenzione della criminalità, ma hanno favorito, in alcune parti del pianeta, anche regimi autoritari nelle loro azioni repressive e di censura<sup>183</sup>.

---

181 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 246-249.

182 Per una ricostruzione della vicenda e per questa interpretazione dei fatti si v. diffusamente A. MANTELERO, *Il futuro regolamento EU sui dati personali e la valenza “politica” del caso Google: ricordare e dimenticare nella digital economy*, cit..

183 Per tutti si v. E. MOROZOV, *L'ingenuità della rete. Il lato oscuro della libertà di internet*, cit., pp. 205-206; D. LYON, *Massima Sicurezza. Sorveglianza e “guerra al terrorismo”*, cit., p. 55 ss.; V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., pp. 246-249. Per le vicende legate al *Datagate* si v. F. PIZZETTI, *Datagate, Prism, caso Snowden: il mondo tra nuova grande guerra cibernetica e controllo globale*, cit..

## 7. I *Big data* e la regolazione pubblica.

Un'altra interessante implicazione giuridica che è stata evidenziata da documenti e letteratura scientifica, anche se ancora superficialmente, e che vale la pena sottolineare è quello inerente l'influenza dell'utilizzo dei *Big data* sulla decisione pubblica e sulla regolazione, intesa con il duplice profilo di scelta di opzioni regolatorie grazie all'utilizzo dei *Big data* e di regolazione del fenomeno dei *Big data*.

I *Big data* possono aprire prospettive nuove e interessanti per la gestione e la cura della cosa pubblica, in quanto possono essere capaci di influire enormemente sulle possibili scelte regolative dei decisori pubblici. Le elaborazioni e le analisi previsionali e probabilistiche prodotte con l'utilizzo di questa grande mole di dati pubblici e privati e la conseguente capacità di osservare la realtà, anche in maniera del tutto differente rispetto ai punti di vista tradizionali possono portare il decisore pubblico a modificare parzialmente o totalmente le proprie scelte regolative o, una volta decisa quale soluzione adottare, a modificare le opzioni regolatorie da poter intraprendere.

I *Big data*, quindi, migliorando la conoscenza della realtà, evidenziando aspetti e prospettive nuove dei fenomeni che vengono analizzati, possono essere un fattore di modifica, anche radicale, delle scelte regolatorie, le quali si possono rivelare obsolete o del tutto tarate su questioni che, in realtà, non erano rilevanti per la risoluzione di determinati problemi o la soddisfazione di determinati bisogni.

Un recente studio dell'OCSE conferma proprio come le grandi potenzialità sottese all'elaborazione dei *Big data* possano migliorare enormemente i processi decisionali e, in particolare la decisione pubblica<sup>184</sup>. In questo documento l'OCSE parla di *data-driven innovation* per intendere appunto «the use of data and analytics to improve or foster new products, processes, organisational methods and markets»<sup>185</sup>. I dati, nella elaborata e dettagliata visione dell'Organizzazione per la cooperazione e per lo sviluppo economico, in un contesto sempre più globalizzato e digitalizzato, diventano motore di crescita e di benessere e, in particolare, diventano uno dei migliori strumenti di governo a disposizione del decisore pubblico. I dati devono essere considerati come una vera e propria infrastruttura della conoscenza pubblica, determinante per migliorare la comprensione della realtà che circonda i governi e le amministrazioni e per adottare le decisioni in modo consapevole. I dati, secondo il documento, attiverebbero un circolo virtuoso per cui, dopo avere raccolto e conservato i dati utili per risolvere una

<sup>184</sup> Il documento è OCSE, *Data-driven innovation. Big data for growth and well-being*, 6 October 2015.

<sup>185</sup> Per una definizione precisa di *data driven innovation* si v. OCSE, *Data-driven innovation. Big data for growth and well-being*, cit., p. 22 e ss.

determinata criticità o per rispondere ad un determinato bisogno (dopo aver datizzato la realtà, per utilizzare un termine già proposto in questo lavoro), i governi e le amministrazioni elaborano i *Big data* con i nuovi strumenti di analisi al fine di creare un fondamento conoscitivo completo su cui basare le proprie decisioni. Decisioni pubbliche che, solo se prese con una base conoscitiva di questo genere, potrebbero favorire sostanzialmente la crescita economica e migliorare il benessere dei cittadini<sup>186</sup>.

È utile sottolineare, inoltre, come il documento ponga delle questioni, inerenti la valorizzazione dei dati e dei risultati delle loro analisi, che interrogano direttamente le discipline pubblicistiche su vari aspetti: dall'attivazione dei fattori abilitanti la formazione di *Big data* e la loro utilizzazione per la decisione pubblica, alla capacità di governare gli effetti più invasivi degli stessi, soprattutto nei confronti dei cittadini, e di trovare un equilibrio tra esigenze di apertura del patrimonio informativo ed necessità di tutela di alcune tipologie di dati.

Per l'OCSE le sfide per il decisore pubblico sono legate essenzialmente all'attivazione dei fattori abilitanti il corretto e proficuo utilizzo dei *Big data*. Innanzitutto la crescita esponenziale della possibilità di produrre dati sulla realtà, grazie allo sviluppo tecnologico e alla diffusione del c.d. *Internet of Things*. Il secondo fattore abilitante è la crescente diffusione delle possibilità di poter utilizzare ed analizzare questi dati. Il documento, come accennavamo in precedenza, considera i dati come un'infrastruttura su cui costruire una nuova modalità decisionale, i quali devono essere i più aperti e accessibili possibili: l'apertura dei dati pubblici (*open data*), la messa a disposizione dei propri dati da parte dei privati (*data philanthropy*) e la maggiore diffusione possibile dei dati scientifici (*data commons*) sono gli elementi imprescindibili per attivare il *data-driven innovation*<sup>187</sup>. Un fattore abilitante, quello della circolazione dei dati, che pone ulteriori implicazioni giuridiche, appunto legate alla ricerca dell'equilibrio tra trasparenza e diffusione dei dati e protezione della privacy e della proprietà intellettuale o industriale.

Il terzo è quello maggiormente utile per comprendere effettivamente l'essenza del *data-driven innovation*: il cambio di paradigma legato alla conoscenza e alla decisione pubblica. Siamo in presenza di un fenomeno che costringe il decisore pubblico, se ne vuole sfruttare a pieno le potenzialità, a cambiare il paradigma legato alla conoscenza pubblica e, quindi, a dividere in maniera sempre più netta le fasi che portano alla decisione finale, in particolare quella della conoscenza dei fenomeni da quella della decisione effettiva.

---

186 OCSE, *Data-driven innovation. Big data for growth and well-being*, cit., p. 35 ss. Così anche G. MISURACA, F. MAREDDU, D. OSIMO, *Policy Making 2.0: Unleashing the power of big data for public governance*, in Springer / Heidelberg Dordrecht, 2014.

187 OCSE, *Data-driven innovation. Big data for growth and well-being*, cit., p. 177 ss.

Bisogna essere capaci di scindere, secondo l'OCSE, in maniera netta i concetti di dato, informazione e conoscenza e conseguentemente il momento della raccolta dei dati da quello dell'estrazione di informazioni dai dati e, ancora, quest'ultimo da quello della formazione della conoscenza che porta alla decisione. Bisogna fare, inoltre, un ulteriore sforzo che permetta di scindere nettamente il momento della formazione della conoscenza da quello della decisione vera e propria<sup>188</sup>.

Anche lo stesso paradigma della decisione pubblica è oggetto di forti mutamenti a causa degli effetti prodotti dall'utilizzo dei *Big data* e che il documento dell'OCSE sembra solo accennare: si sta andando nella direzione di una automatizzazione della decisione pubblica<sup>189</sup>, anche se una maggiore divisione della fase della conoscenza dalla fase decisionale, a parere di chi scrive, responsabilizza maggiormente il decisore pubblico e implica un maggiore utilizzo della sua discrezionalità per prendere una decisione definitiva.

Da questa breve ricostruzione degli aspetti salienti del corposo documento dell'OCSE, si evince come il nuovo paradigma della conoscenza e della decisione, come lo abbiamo ricostruito in queste pagine, coinvolga tutti i livelli della decisione pubblica e tutte le discipline che quei livelli studiano. Dalla decisione prettamente legislativa e, sostanzialmente, dalla decisione su quali opzioni regolatorie e quali scelte regolative adottare, alla decisione amministrativa, che, all'interno dei limiti delineate dalle prime, deve compiere scelte ulteriori e altrettanto determinanti per la risoluzioni di criticità e per la risposta a bisogni sociali e individuali.

Oltre a queste brevi riflessioni sul rilievo giuridico dei *Big data* all'interno delle tematiche inerenti la decisione pubblica - su cui, per quanto riguarda le implicazione amministrative, ritorneremo – il secondo profilo della regolazione pubblica, inteso come tipo di regolazione del fenomeno dei *Big data*, merita di essere approfondito.

Il fenomeno dei *Big data* è un fenomeno che, nonostante sia relativamente recente, è pienamente in espansione, nato in determinati settori ed esploso repentinamente in molti altri ambiti, diffuso nel settore privato e utilizzato anche nel settore pubblico. Un fenomeno, come tutti i fenomeni tecnologici, continuamente in evoluzione, in quanto - grazie al suo costante utilizzo e agli ulteriori avanzamenti tecnologici - affinerà i suoi caratteri, produrrà nuovi effetti e farà nascere nuove problematiche.

Decidere se regolare o meno il fenomeno dei *Big data* e, successivamente, decidere quale scelta regolativa fare, è un'operazione che deve necessariamente tenere presente

188 OCSE, *Data-driven innovation. Big data for growth and well-being*, cit., p. 131 ss.

189 OCSE, *Data-driven innovation. Big data for growth and well-being*, cit., p. 35 ss. In questo senso anche un report commissionato dalla Commissione europea, G. MISURACA, F. MAREDDU, D. OSIMO, *Policy Making 2.0: Unleashing the power of big data for public governance*, in Springer / Heidelberg Dordrecht, 2014.

gli aspetti appena descritti, soprattutto perché il fenomeno produce, come abbiamo potuto vedere in queste pagine, notevoli effetti sulla società, sulla politica, sull'economia, e saranno effetti positivi o negativi anche in ragione delle scelte (o non scelte) regolatorie che si compiono.

Il tema della regolazione dei *Big data* è un tema che ancora non è stato indagato pienamente dalla dottrina, però sembra non solo condivisibile, ma anche condiviso tentare di regolare il fenomeno<sup>190</sup>. Una regolazione che abbia, però, come premessa fondamentale la considerazione che la neutralità della tecnologia rispetto alla società, alla politica e all'economia, non ha nessun fondamento. «Technology is neither good nor bad; nor is it neutral»<sup>191</sup>, diceva la storica prima legge sulla tecnologia di Melvin Kranzberg e, in effetti, la tecnologia resta uno strumento che, come abbiamo evidenziato in precedenza, può avvantaggiare categorie sociali, gruppi di interesse, a scapito di altri gruppi o categorie e la sua regolazione resta determinante anche per rendere lo strumento tecnologico più democratico possibile<sup>192</sup>.

Per la verità regolare i fenomeni tecnologici, come i *Big data*, ci sembra necessario anche perché ci convincono le considerazioni compiute da alcuni studiosi in merito ad un tanto scorretto quanto diffuso determinismo tecnologico che si è venuto a formare soprattutto all'interno del dibattito sulla regolazione di internet, per cui la tecnologia è qualcosa di fondamentalmente ingovernabile o un fenomeno che può essere tendenzialmente subito. Le tecnologie non hanno mai avuto un'evoluzione uniforme ovunque esse sono state utilizzate, ma questa è sempre stata influenzata dalla cultura e dalla società in cui queste vengono utilizzate. La regolazione resta, quindi, il maggiore strumento di governo dei fenomeni sociali e culturali, come d'altronde il fenomeno tecnologico è esso stesso<sup>193</sup>.

Il fenomeno dei *Big data* è un fenomeno fortemente intrusivo, in particolare per il suo impatto sui diritti dei cittadini e per questo motivo necessita di essere regolato, ma è un fenomeno sfuggente. La regolazione dei *Big data* è necessaria, ma ci troviamo di fronte un fenomeno in evoluzione e di difficile ricomposizione che renderà le soluzioni regolatorie che si troveranno parziali e mobili per chissà quanto tempo ancora. Certamente per comprendere come regolarlo dobbiamo partire dall'oggetto della

---

190 In questo senso diffusamente MANTELERO A., *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, cit.; V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit.; E. MOROZOV, *L'ingenuità della rete. Il lato oscuro della libertà di internet*, cit..

191 M. KRANZBERG, *Technology and History: Kranzberg's Laws*, in *Technology and Culture*, 1986, p. 544-560.

192 Per una necessaria regolazione dei *Big data* anche il documento della Casa bianca *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014.

193 E. MOROZOV, *L'ingenuità della rete. Il lato oscuro della libertà di internet*, cit., p. 270 e ss.

regolazione, prima di comprendere chi può (tentare) di regolarlo. La questione ruota attorno alla scelta se bisogna regolare i *Big data*, cioè la produzione di *Big data* o se bisogna regolare l'utilizzo dei *Big data*, l'uso che se ne fa dei *Big data*, magari differenziando la regolazione rispetto all'uso che se ne fa o rispetto a chi utilizza i *Big data*.

Abbiamo avuto modo di vedere nel corso del capitolo di che cosa parliamo quando facciamo riferimento al fenomeno in analisi. Il fenomeno dei *Big data* è caratterizzato, per una parte, da dati *off-line*, dati cioè che non provengono dal *web*, che vengono prodotti da sensori, macchinari, rilevatori e un'altra parte, molto rilevante e più conosciuta, di dati *on-line*, che provengono dal *web*, da Internet. I *Big data*, quindi, hanno una natura differente, con un grado di eterogeneità molto alto. Se i dati *off-line* sono dati la cui produzione potrebbe essere regolata, perché hanno una portata e una circolazione circoscritta, i dati *on-line*, non possono essere regolati con facilità, qualsiasi regolazione si intendesse operare, in particolare una regolazione che mirasse a regolarne la produzione. La difficoltà di regolare i *Big data on-line*, infatti, è simile a quella che si è riscontrata nel regolare internet. I *Big data on-line* hanno una portata transnazionale e tendenzialmente volontaristica, con una possibilità di sviluppo sempre maggiore e con un ruolo rilevante delle multinazionali private che detengono questi dati. Il regolatore pubblico si trova nell'impossibilità di governare larga parte del fenomeno dei *Big data*<sup>194</sup>.

Queste considerazioni ci portano a dire che è complicato ragionare sulla possibilità di regolare la produzione dei *Big data*, ma sarebbe sconveniente farlo anche scindendo i *Big data off-line*, dai *Big data on-line*. La regolazione della produzione di *Big data* porterebbe parzialmente (perché questo discorso già non reggerebbe di fronte alle caratteristiche dei *Big data on-line*) alla limitazione dei rischi che si portano dietro, ma limiterebbe anche le potenzialità del fenomeno, che non vanno assolutamente perse.

Allora l'oggetto della regolazione deve essere, dunque, non tanto la produzione dei *Big data* o i *Big data* in quanto tali, ma l'utilizzo che se ne fa dei *Big data*.

---

194 Per tutti si v. P. MINDUS, *Libertà politica e nuove tecnologie*, in *Ragion pratica*, 44/2015, p. 9 e ss.; In una prospettiva prettamente nazionale si v. G. L. CONTI, *La governance dell'Internet: dalla costituzione della rete alla costituzione nella rete*, in M. NISTICÒ – P. PASSAGLIA (A CURA DI), *Internet e Costituzione*, Torino, 2014, p. 77 e ss.; G. AZZARITI, *Lex digitalis e democrazia*, in M. NISTICÒ – P. PASSAGLIA (A CURA DI), *Internet e Costituzione*, cit., p. 351; T. E. FROSINI, *Internet come ordinamento giuridico*, in M. NISTICÒ – P. PASSAGLIA (A CURA DI), *Internet e Costituzione*, cit., p. 57 e ss.; F. MARCELLI, *Internet fra canale di partecipazione politica e strumenti di controllo. Profilo di diritto internazionale*, in F. MARCELLI, P. MARSOCCI, M. PIETRANGELO (A CURA DI), *La rete internet come spazio di partecipazione politica. Una prospettiva giuridica*, Napoli, 2015, p. 18 e ss.

La regolazione auspicabile è una regolazione che tende a disciplinare (e non limitare) l'utilizzo dei *Big data* e a differenziare la disciplina adeguandola alle esigenze di coloro che utilizzano i *Big data*<sup>195</sup>.

In questo senso l'oggetto della regolazione sono i *Big data* che vengono utilizzati dai privati, nella misura in cui ledono la concorrenza, i diritti legati alla proprietà intellettuale o i diritti alla riservatezza dei dati personali e i *Big data* che vengono utilizzati dal pubblico, in particolare dalle pubbliche amministrazioni durante l'azione amministrativa o nell'attuazione delle politiche di apertura e accessibilità generale ai *Big data*.

### 7. I *Big data* e le pubbliche amministrazioni (segue).

Gli aspetti inerenti la regolazione dei *Big data*, che abbiamo descritto nel paragrafo precedente, ci portano ad affrontare un altro aspetto del rilievo giuridico dei *Big data* che, fino ad ora, non è stato oggetto di particolare attenzione, se non, appunto, a margine dei primi tentativi di affrontare le questioni giuridiche che il tema della regolazione dei *Big data* solleva e delle possibili soluzioni alle criticità e ai rischi che il loro utilizzo comporta<sup>196</sup>. Si tratta delle implicazioni giuridiche che l'elaborazione dei *Big data* e il loro utilizzo nell'azione e nell'organizzazione amministrativa fanno emergere.

Le potenzialità dei *Big data*, in particolare quella di produrre dei risultati analitici originali e predittivi sui fenomeni che vengono analizzati attraverso la loro elaborazione, non possono non interrogare la dottrina amministrativistica sugli effetti che queste nuove possibilità conoscitive possano produrre sull'azione amministrativa concreta, sulle fasi del procedimento, sulla decisione finale, sulle relazioni tra amministrazione e cittadino. Così come i *Big data* diventano una possibilità e anche una

---

195 Si v. V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit.; A. MANTELERO, *Il futuro regolamento EU sui dati personali e la valenza "politica" del caso Google: ricordare e dimenticare nella digital economy*, cit.

196 Qui si fa riferimento alla questione inerente l'accessibilità e la pubblicità dei *Big data* (pubblici e privati) in possesso delle pubbliche amministrazioni. Questo aspetto viene individuato ampiamente come una delle politiche regolatorie più efficaci per frenare il potere informativo che il fenomeno dei *Big data* istituisce in capo ai soggetti pubblici e privati che raccolgono e sono capaci di conservare ed elaborare i *Big data*, valorizzando una maggiore diffusione delle informazioni tra i cittadini. In questo senso V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit., p. 231 e ss.; A. MANTELERO, *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, cit., p. 140. Un tema che coinvolge molti aspetti prettamente amministrativistici come la trasparenza amministrativa, la qualità dei dati detenuti dalle pubbliche amministrazioni e il loro riutilizzo.

possibile fonte di rischio per la decisione pubblica, o meglio per la decisione politica, i *Big data* costituiscono certamente una grande opportunità di miglioramento del processo decisionale e della qualità della decisione amministrativa, ma anche uno strumento da usare con cautela.

Le caratteristiche dei *Big data*, in particolare la loro quantità, la loro eterogeneità, la difficoltà ad conservarli e gestirli, inoltre, pongono questioni non meno importanti rispetto a quelle che abbiamo appena sottolineato, che attengono, però, agli aspetti organizzativi delle pubbliche amministrazioni. Qualora si decidesse di utilizzare i *Big data* nel procedimento amministrativo, bisogna innanzitutto saperli anche organizzare. In questo modo vengono in rilievo quasi tutte le questioni giuridiche più rilevanti che attengono all'organizzazione amministrativa: le modalità di conservazione dei dati, la loro qualità, la loro trasparenza, il loro utilizzo e riutilizzo, la loro circolazione all'interno del sistema amministrativo.

Questioni ancora non indagate a fondo dalla dottrina, ma che sono determinanti per comprendere il fenomeno dei *Big data* e il suo impatto sulle amministrazioni pubbliche e che cominceremo a fare nelle prossime pagine, prendendo in considerazione le prime esperienze pratiche che le amministrazioni pubbliche hanno attuato in rapporto ai *Big data*.

## Capitolo 3

### I *Big data* e le amministrazioni pubbliche

Sommario: 1. Le amministrazioni e i *Big data*: una premessa metodologica.; 2. L'azione amministrativa e i *Big data*: un approccio giuridico.; 2.1. Le procedure di accertamento standardizzato dell'amministrazione finanziaria: un modello per comprendere i *Big data* nel procedimento amministrativo.; 2.2. Il valore giuridico degli strumenti accertativi standardizzati e le dinamiche procedurali che innescano: quale lezione per l'elaborazione dei *Big data* nell'azione amministrativa.; 3. L'organizzazione amministrativa e i *Big data*: i primi tentativi di organizzare i *Big data*.; 3.1. Le esperienze di organizzazione dei *Big data*: ISTAT, INPS e INAIL.; 3.1.1. Le politiche di organizzazione dei *Big data* dell'Istituto nazionale di Statistica (ISTAT).; 3.1.2. Le politiche di organizzazione dei *Big data* dell'Istituto nazionale di previdenza sociale (INPS) e dell'Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro (INAIL).; 3.2. L'estrema valorizzazione del patrimonio informativo come fil rouge delle politiche di organizzazione dei *Big data*.; 4. *Big data*, amministrazioni pubbliche e cittadino: elementi per una nuova relazione cittadino-amministrazione?; 4.1. Il caso dell'epatite C: come l'elaborazione e l'integrazione dei dati da parte dei cittadini influisce sulla decisione pubblica.4.2. L'elaborazione dei *Big data* da parte dei cittadini come nuovo strumento di equilibrio del potere informativo.

#### *1. Le amministrazioni e i Big data: una premessa metodologica.*

Nel capitolo precedente abbiamo visto come l'utilizzo e l'elaborazione dei *Big data* in vari settori economici e sociali abbia cominciato a fare emergere rilevanti questioni giuridiche, alcune volte immediatamente evidenti, altre volte meno, o meglio individuabili dopo una più approfondita riflessione. Questa progressiva emersione delle implicazioni giuridiche dei *Big data*, ha portato, conseguentemente, alla formazione delle prime analisi, più o meno strutturate, sull'impatto dei *Big data* su vari settori e su varie questioni di grande rilievo giuridico<sup>197</sup>.

L'attenzione degli studiosi, i contenuti delle prime analisi non scientifiche del fenomeno e i primi documenti ufficiali delle amministrazioni, si sono concentrate principalmente, e giustamente, sui rischi che i *Big data* possono costituire per i diritti e per le libertà dei cittadini. Nel capitolo precedente abbiamo trattato ampiamente dell'impatto che i *Big data* hanno sulla disciplina della privacy, sull'emersione di un suo profilo individuale e collettivo, sul principio di eguaglianza, sulle libertà fondamentali. Aspetti che hanno

---

197 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit.; D. BOLLIER, *The promise and peril of Big data*, Washington, *The Aspen Institute*, 2010;

ispirato le analisi più compiute del fenomeno e che hanno interrogato anche le amministrazioni preposte alla tutela di quei diritti e di quelle libertà<sup>198</sup>.

Ulteriori aspetti di interesse giuridico sono emersi anche in merito a questioni che parzialmente coinvolgono il cittadino. Basti pensare agli effetti dei *Big data* sulle dinamiche concorrenziali all'interno di determinati mercati, come quelli digitali o dei servizi, sulla disciplina antitrust, su quella della proprietà intellettuale e su quella a tutela del consumatore. Delle riflessioni, che hanno segnalato gli effetti negativi del fenomeno, come la *price discrimination* o la tendenza a monopolizzare i mercati da parte di coloro che raccolgono e utilizzano i *Big data* o, ancora, la difficoltà a disciplinarne gli effetti da parte delle norme in materia di tutela del consumatore e della proprietà intellettuale, ma che hanno anche messo in evidenza le potenzialità dei *Big data* a favore di una migliore concorrenza e selezione dei propri clienti<sup>199</sup>.

Ciò che è emerso con molta evidenza, quindi, sono state analisi giuridiche che cercavano di comprendere come limitare il fenomeno. Meno indagate, come abbiamo potuto constatare, sono le problematiche legate a come utilizzare le potenzialità dei *Big data* e della loro elaborazione.

Questi lavori, anche di taglio giuspubblicista, hanno cominciato a sottolineare come le potenzialità dei *Big data* sono tutte rintracciabili nella loro capacità di modificare il processo decisionale, in particolare quello pubblico, passando da un'analisi su come “difendersi” dai *Big data* ad una che metta in rilievo l'aspetto positivo dei *Big data* e cioè come utilizzarli in maniera compiuta e corretta. Grazie alle nuove possibilità di raccogliere ed elaborare i dati, le scelte regolative e le opzioni regolatorie possono cambiare radicalmente e, se correttamente utilizzate, possono migliorare notevolmente la capacità decisionale delle istituzioni e degli organi politici, in quanto basate su una conoscenza concreta e, possiamo dire, tendenzialmente completa della realtà che si deve regolare<sup>200</sup>.

---

198 European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.; *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, Washington, Maggio 2014; A. MANTELETO – G. VACIAGO, *Data protection in a Big data society. Ideas for a future regulation, Digital investigation*, cit..

199 MCKINSEY, *Big data: the next frontier for innovation, competition and productivity*, 2011; si v. l'intervento del Prof. A. ODLYZKO, *Privacy, price discrimination and seeds of capitalism's destruction*, 28 ottobre 2015, Camera dei deputati organizzato dall'Intergruppo parlamentare sull'innovazione tecnologica; *Executive Office of the President of the U.S.A., Big data and differential pricing*, Washington, Febbraio 2015; M. BOGNI – A. DEFANT, *Big data: diritti IP e problemi della privacy*, cit..

200 OCSE, *Data-driven innovation. Big data for growth and well-being*, cit., p. 35 ss. Così anche G. MISURACA, F. MAREDDU, D. OSIMO, *Policy Making 2.0: Unleashing the power of big data for public governance*, in Springer / Heidelberg Dordrecht, 2014; *Executive Office of the President of the U.S.A., Big data: seizing opportunities, preserving values*, cit.

In questo lavoro, invece, cercheremo di approfondire un aspetto ancora poco indagato dalla dottrina giuridica, in particolare dalla dottrina amministrativistica, e cioè il rapporto tra *Big data* e amministrazioni pubbliche e l'impatto che i primi hanno sull'azione e sull'organizzazione delle seconde.

Se il fenomeno dei *Big data* è un fenomeno recente, la presenza e l'utilizzo dei *Big data* nell'amministrazione è un fenomeno ancora più recente e, ad oggi, sono poche le realtà che hanno cominciato ad approcciarsi al fenomeno dei *Big data* in maniera strutturata e scientificamente rilevante. L'intento di questa parte del lavoro è quello di capire in quali ambiti e come vengono utilizzati dalle amministrazioni pubbliche i *Big data* e quali considerazioni giuridiche si possono trarre da queste esperienze.

Come vedremo nel corso del capitolo, nonostante siano pochi i casi in cui i *Big data*, così come li abbiamo descritti e analizzati nel corso di questo lavoro, vengono utilizzati dalle amministrazioni pubbliche, essi sono emblematici per comprendere quali possono essere gli interrogativi e le problematiche che i *Big data* pongono oggi all'amministrazione.

## 2. *L'azione amministrativa e i Big data: un approccio giuridico.*

Il rapporto tra *Big data* e azione amministrativa e i possibili mutamenti che l'utilizzo dei primi possono apportare alla seconda si evincono dai due risultati conoscitivi principali che le elaborazioni dei *Big data* producono: la maggiore chiarezza e conoscenza della realtà e la possibilità di prevedere l'andamento dei fenomeni che si analizzano.

La raccolta e l'elaborazione dei *Big data* ha dato, osservando le tipologie del loro utilizzo che abbiamo elencato - e potenzialmente potrà dare sempre più all'amministrazione - in primo luogo la possibilità di raccogliere tutte le informazioni possibili sui fatti e sui fenomeni che si vogliono indagare: permette all'amministrazione di datizzare la realtà, che costituisce l'aspetto fondamentale e imprescindibile per il corretto utilizzo dei *Big data*. In secondo luogo, si può affermare che le correlazioni e i rinnovati metodi statistici e matematici di elaborazione dei *Big data* si stanno rivelando come dei nuovi strumenti conoscitivi nelle mani delle pubbliche amministrazioni. Le correlazioni, appunto perché scevre da qualsiasi impostazione preliminare, da qualsiasi ipotesi iniziale, sono capaci di evidenziare relazioni originali e poco intuibili tra le variabili scelte, che mettono i fenomeni che vengono analizzati sotto una luce differente. Le correlazioni, bisogna aggiungere, non vogliono legare quelle variabili con un nesso di causalità, ma sono utili per mettere in evidenza una semplice relazione statisticamente rilevante. L'elaborazione dei *Big data*, poi, ha una forte componente

predittiva: la capacità di poter analizzare una notevole quantità di dati, talmente tanto notevole da abbassare sostanzialmente il rischio statistico - almeno a breve-medio termine - e prevedere probabilisticamente le dinamiche che riguardano un determinato fenomeno.

La migliore e completa conoscenza della realtà mette l'amministrazione nelle condizioni di conoscere davvero i fatti, gli interessi, i bisogni della comunità che governa, in maniera sempre più autonoma e sempre più imparziale. Le analisi probabilistiche e tendenzialmente predittive sui fenomeni e la capacità, attraverso le correlazioni, di legare aspetti in via del tutto originale e molte volte impensabile, dà alle amministrazioni, inoltre, la possibilità di migliorare ulteriormente la conoscenza della realtà, ma anche di velocizzare il processo decisionale e migliorare l'efficacia e l'efficienza della sua azione amministrativa, permettendole di capire molto prima le tendenze dei fenomeni che vuole governare.

Le potenzialità conoscitive dei *Big data* - grazie alle sue caratteristiche, ai suoi strumenti e alla capacità di sapere cogliere, almeno potenzialmente, la realtà nella sua totalità e complessità - ci portano a riflettere su come questa nuova conoscenza influisca sull'azione amministrativa, qui intesa come quell'insieme di atti e di regole che toccano l'aspetto dinamico dell'esercizio della funzione: il momento in cui si passa da una previsione astratta di attività amministrative, compiuta dalla legge, a compimento concreto, effettivo, di quelle attività; dalla predeterminazione di un interesse pubblico astratto, all'individuazione dell'interesse pubblico concreto<sup>201</sup>.

Le domande che ci si possono porre, quindi, sono tendenzialmente due: quando e come vengono utilizzati i dati che hanno le caratteristiche dei *Big data* e gli strumenti che questi dati elaborano; quali sono le implicazioni giuridiche legate al loro utilizzo sull'azione amministrativa.

Per rispondere a questi due quesiti è utile prendere come riferimento esemplificativo alcune tipologie di procedimenti che più di tutti si avvicinano a procedimenti amministrativi che potrebbero utilizzare i *Big data*: stiamo parlando dei procedimenti di accertamento standardizzato che sono svolti, durante la propria funzione di controllo,

---

201 In questo senso si v. F. MERLONI, *Istituzioni di diritto amministrativo*, Torino, 2016, p. 221 e ss.. La dottrina tende sempre più ad utilizzare questo concetto di azione amministrativa in luogo di attività. Si v. per esempio N. PAOLANTONIO, A. POLICE, A. ZITO (a cura di), *La pubblica amministrazione e la sua azione*, Torino, 2005; M.A. SANDULLI (a cura di), *L'azione amministrativa*, Milano, 2005; A. MASSERA, *I principi generali dell'azione amministrativa tra ordinamento nazionale e ordinamento comunitario*, in *Dir. amm.*, 2005; R. TOMEI (a cura di), *La nuova disciplina dell'azione amministrativa*, Padova, 2005; V. CERULLI IRELLI (a cura di), *La disciplina generale dell'azione amministrativa*, Napoli, 2006; G. SCIULLO (a cura di), *Le nuove regole dell'azione amministrativa*, Bologna, 2006; M.P. CHITI, G. PALMA (a cura di), *I principi generali dell'azione amministrativa*, Napoli, 2006; M.A. SANDULLI, *Codice dell'azione amministrativa*, Milano 2011.

dall'amministrazione finanziaria<sup>202</sup>. I motivi di questa scelta sono sostanzialmente tre. Il primo è che l'amministrazione finanziaria, nel proprio lavoro di accertamento, utilizza una quantità di dati sui contribuenti enorme e molto eterogenei, i quali compongono un insieme di dati che, per caratteristiche, è molto simile a quelle dei *Big data*, le quali possono essere semplificate nelle c.d. 3 V (Volume, Velocità, Varietà). L'amministrazione finanziaria, infatti, per i propri fini istituzionali, non solo possiede i dati propri - come quelli che i contribuenti forniscono volontariamente all'Agenzia delle entrate per il pagamento delle imposte o come i dati catastali, oramai pienamente in possesso dell'amministrazione finanziaria<sup>203</sup> - ma ha la possibilità di accedere ad una serie di altri dati in possesso di altri enti pubblici o altri soggetti, terzi rispetto al rapporto tra amministrazione e cittadino<sup>204</sup>. L'amministrazione finanziaria, infatti, può accedere ai dati bancari, ai dati in possesso delle altre amministrazioni dello Stato, degli enti locali e regionali, degli enti pubblici non economici, degli enti assicurativi pubblici e privati, delle magistrature e degli organi ispettivi e di vigilanza, dei pubblici ufficiali generalmente intesi (come notai, conservatori dei registri immobiliari, etc.). Essa, inoltre, può accedere anche ai dati riguardanti i contribuenti in possesso di soggetti terzi che sono tenuti a tenere una contabilità dei dati (clienti, fornitori, amministratori di condominio etc.)<sup>205</sup>. L'amministrazione finanziaria, quindi, ha la possibilità di integrare più insieme di dati, i quali possono anche avere un grado di eterogeneità molto elevato.

Il secondo motivo riguarda i metodi di accertamento che vengono utilizzati nei procedimenti di accertamento standardizzato condotti dall'amministrazione finanziaria, i quali sono sostanzialmente delle elaborazioni statistiche e matematiche simili alle correlazioni dei *Big data*, anche se vale la pena sottolinearne le profonde differenze. Terzo e ultimo motivo è legato alle dinamiche procedurali che l'utilizzo di strumenti conoscitivi di questo genere producono e la valenza giuridica esterna che queste elaborazioni hanno.

Motivi che ci portano a considerare questa tipologia di procedimenti di accertamento come un'ottima base per comprendere quali potranno essere le implicazioni giuridiche

202 Per una generale classificazione dei procedimenti di accertamento tributario si v. I. MANZONI – G. VANZ, *Il diritto tributario*, Torino, 2008; G. FALSITTA, *Manuale di diritto tributario. Parte generale*, Padova, 2012; P. ROSSI, *I metodi di accertamento*, in A. FANTOZZI (a cura di), *Diritto tributario*, IV ed., Torino, 2013.

203 Vale la pena ricordare come con l'art. 23-quater del d.l. n. 95/2012 l'Agenzia del territorio fosse stata soppressa e le sue funzioni e tutti gli strumenti utili allo svolgimento delle sue funzioni (compresa la gestione dei dati catastali) siano nella disponibilità dell'Agenzia delle entrate. Per approfondire si v. G. MAZZANTINI, *Il decreto "spending review"*, in *Gior. dir. amm.*, 2012, fasc. 12, p. 1161.

204 Si pensi alla vicenda dell'utilizzo delle medie ISTAT per la delineazione degli elementi del redditometro che ha visto il Garante per la protezione dei dati personali esprimersi in maniera negativa nel 2013.

205 Per una panoramica generale sull'istruttoria e sui controlli dell'amministrazione finanziaria si v. A. MARCHESELLI, *Accertamenti tributari e difesa del contribuente. Poteri e diritti nelle procedure fiscali.*, Milano, 2010, p. 41 ss..

sulle dinamiche procedurali qualora i *Big data* fossero utilizzati costantemente dalle amministrazioni pubbliche per prendere delle decisioni in tutti i settori.

*2.1. Le procedure di accertamento standardizzato dell'amministrazione finanziaria: un modello per comprendere i Big data nel procedimento amministrativo.*

È utile allora ricostruire brevemente a quale tipologie di procedimento di accertamento ci riferiamo quando parliamo di accertamento tributario standardizzato.

Nel diritto tributario ci sono due accezioni di accertamento, una più ampia ed una più ristretta<sup>206</sup>. Ai nostri fini è utile riprendere soltanto la seconda accezione, quella più ristretta, in quanto essa fa riferimento a quel «complesso di attività poste in essere dai soggetti titolari di potestà impositiva (amministrazione finanziaria, Agenzie fiscali, enti territoriali) e oggi caratterizzate, essenzialmente, dalla loro conformazione come controlli sugli adempimenti dei soggetti passivi (i contribuenti)»<sup>207</sup>, escludendo tutti gli adempimenti e le attività che il contribuente è tenuto a fare in attuazione delle norme tributarie.

Le attività di controllo e di accertamento tributario sono attività vincolate, che costituiscono la fase istruttoria di un unico e unitario procedimento che porta all'adozione di atti autoritativi di accertamento<sup>208</sup>. Queste hanno una duplice finalità: da un lato, la contestazione di illeciti e il conseguente recupero della materia imponibile sottratta ad imposizione, quindi, una prima finalità riconducibile ad attività di “polizia amministrativa”, e l'acquisizione di elementi di fatto, dati e notizie fiscalmente rilevanti, necessarie all'amministrazione per conoscere la realtà in cui si trova ad operare e indirizzare le proprie attività in modo efficiente ed imparziale<sup>209</sup>.

---

206 Secondo la dottrina tributaria per accertamento in senso più ampio si intende ricomprendere tutte quelle «attività proprie dell'attuazione della norma tributaria finalizzate a stabilire che un presupposto si sia verificato, a qualificarlo, a misurarlo mediante la determinazione della base imponibile, a valutare l'applicabilità delle deduzioni, detrazioni e altre agevolazioni». Come è possibile osservare nell'accezione ampia di accertamento rientrerebbero anche le attività che il contribuente compie per adempiere al suo dovere di solidarietà e che esulano dalla nostra trattazione. Così in M. BASILAVECCHIA, *Accertamento (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, Milano, 2006, p. 45 e ss..

207 Così M. BASILAVECCHIA, *Accertamento (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, cit., p. 47 e ss.. Le parentesi sono le nostre.

208 G. VANZ, *Controlli amministrativi (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, cit., p. 1438 e ss.; S. LA ROSA, *Principi di diritto tributario*, Torino, 2004, p. 115-116; R. SCHIAVOLIN, *Poteri istruttori dell'amministrazione finanziaria*, in *Dig. Disc. Priv. Sez. Comm.*, XI, Torino, 1995, p. 194 e ss.

Per compiere le attività di controllo, che sono fondamentalmente attività istruttorie e conoscitive, al fine di perseguire le attività appena elencate, la legge affida all'amministrazione finanziaria poteri di accertamento, che vanno dalla richiesta di informazioni al contribuente oppure ordini di trasmissione o, ancora, di esibizione di documenti, all'acquisizione di dati bancari, sino a strumenti molto più penetranti come l'accesso diretto presso l'azienda, lo studio professionale di riferimento del contribuente o, nei casi più gravi e previa autorizzazione giudiziaria, presso l'abitazione del contribuente interessato<sup>210</sup>.

L'amministrazione finanziaria, però, nonostante sia vincolata sulle vicende che riguardano il fatto imponibile, in particolare nella sua ricostruzione e nella sua valutazione, e, in generale, nella sua attività di controllo, ha un maggiore spazio discrezionale, anche se fondamentalmente tecnico<sup>211</sup>, in merito alla scelta del contribuente da controllare e ai metodi di accertamento che possono essere utilizzati<sup>212</sup>. Per metodo di accertamento si intendono «le tecniche con le quali l'autodeterminazione del contribuente (in particolare: la dichiarazione) viene controllata e, se del caso, rettificata o sostituita, e gli elementi probatori che devono porsi a sostegno di tali tecniche»<sup>213</sup>.

Quando parliamo di accertamento standardizzato parliamo fondamentalmente di «una tipologia di metodi di accertamento, i quali – sebbene siano caratterizzati da ambiti applicativi e logiche di funzionamento diversi – trovano nel ricollegarsi a condizioni di normalità economica un tratto omogenizzante»<sup>214</sup>: stiamo parlando dei parametri e degli studi di settore, da un lato, e dell'accertamento sintetico basato su indici nominati (c.d. redditometro), dall'altro.

---

209 G. VANZ, *Controlli amministrativi (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, cit., p. 1439 e ss., ma più in generale F. LEVI, *L'attività conoscitiva della pubblica amministrazione*, Torino, 1967.

210 L. SALVINI, *La partecipazione del privato all'accertamento*, Padova, 1990, p. 67 e ss.; I. MANZONI, *Potere di accertamento e tutela del contribuente*, Milano, 1993, p. 213 e ss.; A. VIOTTO, *I poteri di indagine dell'amministrazione finanziaria*, Milano, 2002, p. 184 e ss.; A. MARCHESSELLI, *Accertamenti tributari e difesa del contribuente. Poteri e diritti nelle procedure fiscali.*, cit..

211 Sul punto è interessante rilevare come la dottrina tributaristica sia tendenzialmente unanime nell'affermare come nei procedimenti tributari non ci sia mai discrezionalità se non una discrezionalità tecnica in quanto non viene mai garantita all'amministrazione finanziaria una reale possibilità di ponderare l'interesse primario (della pronta e perequata percezione dei tributi) con eventuali interessi secondari (sia pubblici, sia privati). In questo senso F. GALLO, *Discrezionalità dell'accertamento tributario e sindacabilità delle scelte dell'ufficio*, *Riv. Dir. Fin.*, 1992, p. 655 e ss.; L. PERRONE, *Discrezionalità amministrativa (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, cit., p. 2003 e ss.

212 G. GAFFURI, *I metodi di accertamento fiscale e le regole costituzionali*, in *Studi in onore di Uckmar*, Padova, 1997.

213 M. BASILAVECCHIA, *Accertamento (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, cit., p. 53 e ss.

214 E. DELLA VALLE, *Note minime in tema di accertamenti standardizzati*, in *Rass. Trib.*, 2014, fasc. 4, p. 695 e ss..

Entrambi i metodi di accertamento sono legati ad un modello di normalità economica che ha come punto fondamentale una predeterminazione normativa che fornisce uno *standard* capace di orientare gli uffici dell'amministrazione finanziaria nella verifica della "congruità" dei ricavi e dei compensi dichiarati da soggetti che sono obbligati dalla legge a tenere delle scritture contabili, nel caso dei parametri e degli studi di settore, e nella ricostruzione del reddito complessivo delle persone fisiche muovendo dalle spese da loro effettuate e, più generalmente, dal loro tenore di vita, nel caso del c.d. redditometro<sup>215</sup>.

Una nozione, quella di accertamenti standardizzati che, nonostante sia stata per molto tempo minoritaria in dottrina e un approdo recente nella giurisprudenza tributaria<sup>216</sup>, ci è molto utile per la nostra analisi, soprattutto per la sua capacità di racchiudere in una stessa tipologia di metodi di accertamento quelli che, più di tutti, costituiscono degli interessanti strumenti di elaborazione dei dati nel procedimento amministrativo, che si avvicinano moltissimo ad un possibile utilizzo delle elaborazioni dei *Big data* nell'azione amministrativa.

Gli studi di settore e il redditometro, infatti, sono strumenti di accertamento che utilizzano funzioni matematiche e inferenze statistiche per presumere i costi e i ricavi di un'impresa, il primo, e il reddito complessivo di una persona fisica, il secondo. Per comprendere meglio i due strumenti è utile allora cercare, in sommi capi, di chiarire come funzionano.

Gli studi di settore, ricordiamo, sono utili all'amministrazione finanziaria per individuare le condizioni effettive di operatività di imprese e professionisti e a determinare i ricavi e i compensi che, con ragionevole probabilità, possono essere loro attribuiti. Il legislatore, con l'art. 10 della l. n. 146/1998, ha introdotto una presunzione che attribuisce al contribuente ricavi e compensi sulla base di alcuni fatti (interni o esterni all'azienda) che denotano una potenziale capacità di produrli.

Per determinare i ricavi e i compensi vengono definiti dei *cluster*, cioè dei gruppi omogenei di imprese. L'omogeneità si deduce da alcuni fattori (modelli organizzativi adottati, dalle aree di mercato cui l'impresa si rivolge etc.). Per ciascun *cluster* viene definita una funzione matematica che consenta di dedurre i ricavi dai dati contabili e strutturati dell'impresa. Il contribuente presenta la propria dichiarazione, attraverso il *software* Ge.Ri.Co. (Gestione dei Ricavi o Compensi) il quale consente l'individuazione

---

215 Ibidem.

216 A. MARCHESELLI, *Le presunzioni nel diritto tributario: dalle stime agli studi di settore*, Torino, 2008, p. 274 e ss.. Di rilievo anche la posizione della giurisprudenza tributaria come in Cass., SS.UU., 18.12.2009, n. 26638, con nota di M. BASILAVECCHIA, *Accertamento e studi di settore: soluzione finale*, in *Riv. Giur. Trib.*, 2010, 205 e ss. e in Cass., SS.UU., 18.12.2009, n. 26635, con nota di A. MARCHESELLI, *La Sezione Unite sulla natura presuntiva degli studi di settore*, in *Corr. Trib.*, 2010, p. 255.

del proprio *cluster* di appartenenza e del volume dei ricavi che dovrebbe corrispondere allo stesso. Il programma, inoltre, consente anche di valutare la congruità dei propri ricavi - cioè se rientrano o meno nell'intervallo di “confidenza parametrica” indicato nel *cluster* - e la coerenza di alcuni indicatori economici – come per esempio la produttività per addetto, la resa oraria, il tasso di rotazione del magazzino - con i valori massimi e minimi fissati per ciascun *cluster*. Dopo di che e a seconda dei risultati dell'elaborazione dei dati da parte del *software* ci possono essere o meno, a seconda del discostamento della singola impresa dal proprio *cluster*, le ulteriori fasi del procedimento tributario di accertamento<sup>217</sup>.

Allo stesso modo, il redditometro, costituito da un elenco di alcuni beni o servizi la cui disponibilità è considerata un elemento indicativo della capacità contributiva, è il migliore strumento di accertamento sintetico, quindi presuntivo, per il controllo di massa delle dichiarazioni delle persone fisiche. L'art. 38 del D.P.R. n. 600/1973 autorizza il Ministero dell'Economia e delle Finanze ad individuare, attraverso un proprio regolamento, alcuni fatti-indice - ad esempio, la disponibilità di imbarcazioni da diporto, le spese per collaborazioni domestiche etc. - e a collegarle ad alcuni coefficienti. L'amministrazione finanziaria, nella sua concreta azione amministrativa di controllo, deve rilevare la sussistenza dei fatti-indice e, successivamente, applicare i coefficienti fissati con il suddetto regolamento. In questo modo l'amministrazione finanziaria potrà determinare il reddito complessivo netto presunto del contribuente sotto controllo. Se il reddito complessivo netto presunto è superiore per almeno un quarto a quello dichiarato e se questo discostamento è riscontrabile per due anni di fila, l'amministrazione può procedere con la rettifica del reddito<sup>218</sup>.

La breve descrizione di questi due metodi di accertamento, gli studi di settore e il redditometro, ci è utile per comprendere come questi due strumenti possano essere avvicinati, ai fini dell'analisi che stiamo facendo in questo lavoro, all'elaborazione dei *Big data*.

Sia gli studi di settore che il redditometro sono tecniche di accertamento, utilizzate dall'amministrazione finanziaria durante l'attività di controllo, che costituisce la fase dell'istruttoria procedimentale, che utilizzano delle funzioni matematiche e statistiche per le finalità per cui è stata preposta: sia di recupero della base imponibile e di

---

217 Sugli studi di settore si v. D. DI SABATO, *La congruità del contribuente e la coerenza del sistema*, in *Riv. Dir. Dell'impr.*, 2014, fasc. 2, pag. 241 – 254; G. GIRELLI, *Gli studi di settore quale strumento "multifunzionale" tra dichiarazione, accertamento e processo tributario*, in *Riv. Dir. Trib.*, 2012, fasc. 7-8, pag. 721 – 757; M. VERSIGLIONI, *Prova e studi di settore*, Milano, 2007.

218 Sul redditometro si v. M. BEGHIN, *Profili sistematici e questioni aperte in tema di accertamento «sintetico» e «sintetico-redditometrico»*, in *Riv. Dir. Trib.*, 2010, I, 717 e ss.; A. MARCHESELLI, *Le presunzioni nel diritto tributario: dalle stime agli studi di settore*, cit.; F. BASTIONI FERRARA, *I principi della riforma tributaria: accertamento sintetico e redditometro*, in *Riv. Dir. Trib.*, 1994, I, 705 e ss.

deterrenza verso comportamenti evasivi dell'imposizione tributaria, sia di conoscenza della realtà concreta di azione amministrativa. Delle funzioni matematiche e statistiche che forniscono una ricostruzione probabilistica e descrittiva dei fenomeni che analizzano: una fotografia di massima del settore economico di appartenenza, nel caso degli studi di settore, o una ricostruzione probabilistica del reddito complessivo, nel caso del redditometro. Anche le elaborazioni dei *Big data* possono essere utilizzate per finalità descrittive, ma, a differenza delle tecniche di accertamento in analisi, utilizzano strumenti statistici molto più immediati, come le correlazioni non lineari, i cui risultati possono avere una consistenza scientifica meno solida, ma posseggono una propensione innovativa molto più elevata, sia in termini descrittivi – si pensi alla possibilità di rilevare aspetti originali e differenti dei fenomeni in analisi – sia in termini predittivi, grazie alla possibilità di prevedere probabilisticamente il *trend* evolutivo del fenomeno analizzato.

Nonostante alcune differenze, anche sostanziali delle tecniche utilizzate e delle caratteristiche dei risultati prodotti, le tecniche di accertamento standardizzato hanno un altro elemento in comune con le elaborazioni dei *Big data*, che le rende un oggetto di analisi fondamentale per i nostri obiettivi. Oltre all'utilizzo di funzioni statistiche e matematiche, di elaborazioni probabilistiche all'interno dell'azione amministrativa, sia gli strumenti accertativi standardizzati sia l'elaborazione dei *Big data*, costruiscono uno schema di regolarità da cui è possibile rilevare degli scostamenti individuali da determinati parametri, che possono avere, a seconda dell'utilizzo che se ne fa, determinati effetti sui cittadini.

Nel caso degli accertamenti standardizzati, l'utilizzo degli strumenti in analisi, appunto grazie ad uno schema regolarità-scostamento dai parametri di normalità economica, che sia il redditometro che gli studi di settore utilizzano, porta all'individuazione dei soggetti (imprese o persone fisiche) che si discostano da quei parametri: un discostamento che, in quanto indicatore di una possibile evasione o irregolarità, potrebbe attivare i penetranti poteri di accertamento in capo all'amministrazione finanziaria e portare all'adozione di un atto di accertamento nei confronti del soggetto interessato.

Allo stesso modo l'utilizzo delle elaborazioni dei *Big data*, come abbiamo potuto vedere nei capitoli precedenti, in particolare quando vengono utilizzati in funzione di controllo e di sicurezza e prevenzione della criminalità, produce lo stesso schema di regolarità-scostamento da cui vengono fatte discendere valutazioni ed effetti sui soggetti interessati di non poco momento.

Osservare, allora, quale valore giuridico hanno questi strumenti e quali dinamiche procedurali innescano, può essere utile per trovare delle indicazioni problematiche sull'utilizzo delle elaborazioni dei *Big data* nell'azione amministrativa.

*2.2. Il valore giuridico degli strumenti accertativi standardizzati e le dinamiche procedurali che innescano: quale lezione per l'elaborazione dei Big data nell'azione amministrativa.*

Nel paragrafo precedente abbiamo potuto vedere come il funzionamento degli strumenti di accertamento standardizzato sia molto simile all'utilizzo delle elaborazioni dei *Big data* durante l'azione amministrativa. Entrambi gli strumenti utilizzano dati molto differenti tra di loro e in grande quantità; utilizzano funzioni matematiche e statistiche per produrre determinati risultati; questi risultati vengono utilizzati, vedremo in che termini, per valutare e per prendere decisioni; queste decisioni possono avere degli effetti, anche molto penetranti sui soggetti a cui sono rivolte.

La prima questione rilevante è ricostruire quale valore giuridico hanno gli strumenti di accertamento standardizzato come gli studi di settore e il redditometro.

Nonostante non ci siano unanimi interpretazioni in merito nella dottrina tributaristica<sup>219</sup>, possiamo dire che entrambi gli strumenti sono considerati come presunzioni semplici.

Le presunzioni semplici sono disciplinate dagli artt. 2727 e 2729 del Codice civile.

L'art. 2727 c.c. stabilisce che «le presunzioni sono le conseguenze che la legge o il giudice trae da un fatto noto per risalire ad un fatto ignoto». Il codice distingue tra presunzioni legali, disciplinate dall'art. 2728, e presunzioni semplici, ex. art. 2729 del c.c.. La differenza tra le due presunzioni è fortemente legata alla questione probatoria. Mentre le presunzioni legali sono stabilite dalla legge e dunque dispensano coloro a favore delle quali sono state stabilite dal dover fornire una prova, le presunzioni semplici non sono stabilite dalla legge e non costituiscono una prova, in quanto vengono rilasciate alla valutazione del giudice che potrà ammetterle solo se gravi (rilevanti), precise (con una interpretazione univoca) e concordanti (coerenti con il contesto). Ne discende che, nel caso di una presunzione semplice, la parte che intende trarne un vantaggio avrà l'onere di dimostrarne il fondamento; in pratica, nelle dinamiche tributarie, la parte che intende avvalersene, di fatto l'amministrazione finanziaria, avrà l'onere di provare che gli elementi presuntivi posti a base della pretesa impositiva hanno i caratteri di gravità, precisione e concordanza.

---

<sup>219</sup> Considerano presunzioni legali relative tutti gli strumenti di accertamento standardizzato I. MANZONI – G. VANZ, *Il diritto tributario*, cit.; G. FALSITTA, *Manuale di diritto tributario. Parte generale*, cit.; P. ROSSI, *I metodi di accertamento*, in A. FANTOZZI (a cura di), *Diritto tributario*, cit.; *Contra* E. DELLA VALLE, *Note minime in tema di accertamenti standardizzati*, cit.; A. MARCHESELLI, *Le presunzioni nel diritto tributario: dalle stime agli studi di settore*, cit..

In realtà, come avevamo accennato in precedenza la riconduzione alla medesima natura di presunzioni semplici del redditometro e degli studi di settore sono asserzioni giurisprudenziali recenti e oscillanti<sup>220</sup>.

Emblematica, però, ci pare una sentenza della Cassazione Sezione Unite del 2009 in cui si afferma che «una lettura costituzionalmente orientata delle disposizioni relative alla procedura di accertamento mediante l'applicazione dei parametri impone di attribuire a quest'ultimi una natura meramente presuntiva, in quanto, a ben vedere, essi non costituiscono un fatto noto e certo, capace di rivelare con rilevante probabilità il presunto reddito del contribuente, ma solo una estrapolazione statistica di una pluralità di dati settoriali elaborati sulla base dell'analisi delle dichiarazioni di un campione di contribuenti»<sup>221</sup>. Considerazioni della giurisprudenza che, fatte in merito alla natura dei parametri e degli studi di settore, sono state poi estese anche al redditometro in un'altra sentenza della Cassazione Sezioni Unite del 2012<sup>222</sup>.

Tra l'altro bisogna sottolineare come, sia la sentenza della Cassazione appena citata, sia alcuni studiosi, hanno messo in evidenza come affinché una presunzione possa essere considerata semplice è necessario che il fatto su cui si basa l'indizio sia certo. Di tale certezza non v'è traccia in schemi, come quelli in analisi, che, per definizione, guardano a condizioni di normalità economica e non a condizioni concrete della singola impresa o del singolo contribuente<sup>223</sup>.

Il riferimento alla presunzione semplice sembra doversi leggere, allora, non in maniera assoluta, ma come un modo strumentale al riconoscimento dell'insufficienza del dato matematico-statistico, risultato dell'utilizzo degli studi di settore e del redditometro, per fondare l'avviso di accertamento. Il discostamento rispetto all'esito dell'elaborazione statistica, infatti, proprio perché riferito ad una condizione di supposta normalità economica, costituisce, in una catena argomentativa più ampia, soltanto uno degli elementi a sostegno della pretesa erariale. Il che implica, quale ulteriore e logico corollario, che l'amministrazione finanziaria debba raccogliere ulteriori elementi su cui

---

220 Nel senso di considerare gli strumenti di accertamento standardizzato come presunzioni semplici si v. Cass., sez. trib., 20.12.2012, n. 23554, con nota di M. BEGHIN, *Il redditometro e gli altri accertamenti per "standard" nelle maglie della presunzione semplice*, in *Corr. Trib.*, 2013, p. 2042. Nello stesso senso, più di recente, ved. Cass., sez. VI, ord. 6.2.2013, n. 2806, reperibile con nota di L. LOVECCHIO, *Spese per incrementi patrimoniali imputabili secondo il principio della quota risparmio anche nel "nuovo redditometro"*, in *Riv. giur. trib.*, 2013, p. 333. Per un'interpretazione differente, che vede i parametri redditometrici come presunzioni legali relative si v. Cass., sez. trib., 30.9.2005, n. 19252; Cass., sez. trib., 23.7.2007, n. 16284; Cass., sez. trib., 19.3.2014, n. 6396

221 Cass., SS.UU., 18.12.2009, n. 26638, con nota di M. BASILAVECCHIA, *Accertamento e studi di settore: soluzione finale*, cit. e in Cass., SS.UU., 18.12.2009, n. 26635, con nota di A. MARCHESELLI, *La Sezione Unite sulla natura presuntiva degli studi di settore*, cit.

222 Cass., sez. trib., 20.12.2012, n. 23554, con nota di M. BEGHIN, *Il redditometro e gli altri accertamenti per "standard" nelle maglie della presunzione semplice*, cit..

223 E. DELLA VALLE, *Note minime in tema di accertamenti standardizzati*, cit..

fondare l'avviso di accertamento e, in giudizio, che il giudice, nel fondare la sua decisione, dovrà dar conto di aver riscontrato che il nesso inferenziale tra il fatto (solo statisticamente) noto ed il fatto ignoto è rilevante (gravità), univoco (precisione) e coerente con il contesto (concordanza)<sup>224</sup>.

Gli studi di settore e il redditometro, quindi, possono essere strumenti istruttori molto utili per una prima valutazione di massa della correttezza delle dichiarazioni, ma i loro risultati non possono costituire da soli un presupposto per l'avviso di accertamento.

Il vero aspetto problematico della questione sarebbe appunto il legame tra la natura degli accertamenti standardizzati – presunzioni semplici - e la loro capacità di attivare poteri di controllo molto penetranti da parte dell'amministrazione finanziaria.

I poteri di controllo dell'amministrazione finanziaria si esplicano imponendo al soggetto interessato un *facere* (ad esempio comparire di persona per dare delle informazioni), un *dare* (ad esempio trasmettere dei documenti o delle informazioni) oppure, nei casi più gravi, un *patis* (ad esempio subire un accesso ispettivo), con una ricaduta evidente sulle libertà individuali del privato. Le attività di controllo hanno, quindi, una diretta incidenza sulla libertà personale, sulla libertà di domicilio, su quella di comunicazione, sul segreto professionale e bancario, in generale, sul diritto alla riservatezza; tutte libertà garantite da disposizioni costituzionali, come gli artt. 13-14-15 della Costituzione, o da norme internazionali come l'art. 8 della Convenzione europea per la salvaguardia dei diritti dell'uomo e delle libertà fondamentali<sup>225</sup>.

Per questi motivi il procedimento tributario, in particolare l'attività di controllo, è un procedimento fortemente ancorato al principio di legalità e, di conseguenza fortemente tipizzato, in cui i poteri che abbiamo elencato precedentemente sono attivabili per specifiche e determinate finalità e in specifiche e determinate situazioni. Oltre al principio di legalità, anche i principi generali che reggono il procedimento amministrativo, come il principio di ragionevolezza e il principio di proporzionalità dell'intervento, caratterizzano fortemente il procedimento tributario, in particolare le attività di controllo che utilizzano strumenti di accertamento standardizzato<sup>226</sup>.

Nonostante i principi appena elencati reggano il procedimento amministrativo e tributario almeno dall'approvazione della legge n. 241/1990<sup>227</sup>, la strutturazione e

---

224 Ibidem.

225 Per questa ricostruzione si v. G. VANZ, *Controlli amministrativi (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, cit., p. 1440.

226 Sul procedimento tributario si v. L. SALVINI, *Procedimento amministrativo (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, cit., p. 4531; A. MARCHESELLI, *Il "giusto procedimento" tributario. Principi e discipline*, Padova, 2012.

227 Sui principi di ragionevolezza e di proporzionalità si v. A. SANDULLI, *La proporzionalità dell'azione amministrativa*, Padova, 1998; D.U. GALETTA, *Proporzionalità e sindacato giurisdizionale nel diritto amministrativo*, Milano, 1998; J. SCHWARZE, *The Principle of*

l'utilizzo di questi strumenti di accertamento possono inficiare alcuni di essi come, ad esempio, il principio di proporzionalità e di ragionevolezza dell'azione amministrativa. In questo senso ci pare molto utile riportare le considerazioni che il Garante della privacy ha fatto nel novembre del 2013 esprimendo forti perplessità sulla plausibilità del modo in cui il redditometro è costruito, sulla ragionevole attendibilità dei suoi risultati e, più in generale, sulla proporzione tra strumenti accertativi e diritti dei contribuenti<sup>228</sup>.

Per quanto riguarda i primi due punti, inerenti la credibilità dello strumento e dei suoi risultati, il Garante ha sostanzialmente imposto l'espunzione di tutte le c.d. spese correnti imputabili mediante riferimento alle "spese medie ISTAT" dalla determinazione presuntiva del reddito complessivo del contribuente<sup>229</sup>, riducendo drasticamente il peso degli elementi "stimati" in base a dati di normalità economica nel redditometro, in quanto rendevano lo strumento di accertamento eccessivamente presuntivo. La determinazione del reddito complessivo risulta ora basata essenzialmente sulle "spese certe" e sulle "spese per elementi certi", da un lato, e sugli incrementi patrimoniali e sulla quota di risparmio<sup>230</sup>, anche se resta, per la dottrina, comunque una determinazione presuntiva e una presunzione semplice<sup>231</sup>.

Questo ridimensionamento dell'incertezza degli elementi costituenti il redditometro rientra nella più generale riflessione del Garante sulla proporzione tra strumenti accertativi e diritti dei contribuenti, che conferma l'impostazione di una giurisprudenza molto minoritaria<sup>232</sup>. Il Garante sottolinea come l'utilizzo del redditometro e il

---

*Proportionality and the Principle of impartiality in European Administrative Law*, in *Riv. trim. dir. Pubbl.*, 2003; S. COGNETTI, *Principio di proporzionalità. Profili di teoria generale e di analisi sistematica*, Torino, 2011; F. MERUSI, *Ragionevolezza e discrezionalità amministrativa*, Napoli 2011; M. RENNA, F. SATTA, *Studi sui principi del diritto amministrativo*, Milano, 2012; G. ROSSI, *L'attualità di G.D. Romagnosi nell'eclissi dello statalismo. Considerazioni sul passato e sul futuro del diritto amministrativo*, in *Dir. pubbl.*, 2012.

228 Parere del Garante per la protezione dei dati personali, *Redditometro: le garanzie dell'Autorità a seguito della verifica preliminare sul trattamento di dati personali effettuato dall'Agenzia delle entrate* del 21 novembre 2013. Per i primi commenti M. BASILAVECCHIA, *Privacy e accertamento sintetico: primi segnali di riequilibrio*, in *Corr. Trib.*, 2014, p. 9 ss.; A. MARCHESELLI, *Redditometro e diritti fondamentali: da Garante e giurisprudenza estera un «assist» ai giudici tributari italiani*, in *Corr. Trib.*, 2014, p. 14 e ss.

229 In realtà è stato evidenziato come le spese per elementi certi possono essere determinabili anche facendo riferimento alle spese medie ISTAT, quando non sono determinabili in modo obiettivo. Così in A. CONTRINO – A. MARCHESELLI, *Il «redditometro 2.0» tra esigenze di privacy, efficienza dell'accertamento e tutela del contribuente*, in *Dir. E Prat. Trib.*, fasc. 4, 2014, p. 10679 e ss.

230 In questo senso la Circolare dell'Agenzia delle Entrate, 11 marzo 2014, n. 6/E che recepisce le osservazioni del Garante e integra la Circolare dell'Agenzia delle Entrate, 31 luglio 2013, n. 24/E di commento generale alla disciplina del redditometro.

231 Le considerazioni del Garante hanno riaperto il dibattito sulla natura di presunzione semplice o di presunzione legale relativa del redditometro. Per tutti si v. A. CONTRINO – A. MARCHESELLI, *Il «redditometro 2.0» tra esigenze di privacy, efficienza dell'accertamento e tutela del contribuente*, cit..

232 Trib. Napoli, Sez. di Pozzuoli, Ord. 21 febbraio 2013, con nota di S. GIORDANO, *Potenziati e irreparabili danni alla libertà personale con il "nuovo redditometro"*, in *GT – Riv. Giur. Trib.*, 2013, fasc. 4, pag. 356 – 362; Comm. Trib. Prov. - Reggio Emilia, sez. II, 18 aprile 2013, n. 74, con nota di

conseguente ingresso nella vita privata del contribuente dovrebbe essere legato ad elementi stabiliti dalla legge (come ad esempio indizi di frode o comunque un tenore di vita che faccia ragionevolmente sospettare la presenza di comportamenti evasivi). Una considerazione che, secondo la dottrina più accorta, si aggancerebbe alle migliori giurisprudenze estere e internazionali sul punto e che evidenzerebbe la sproporzione dell'utilizzo di questi strumenti nel nostro ordinamento, in quanto l'utilizzo di strumenti di accertamento standardizzato non sono ricollegati a nessun elemento particolare, ma possono essere sempre utilizzati per attività di controllo<sup>233</sup>.

Le considerazioni del Garante della protezione dei dati personali sul redditometro, che per alcuni studiosi dovrebbero essere estesi anche agli studi di settore<sup>234</sup>, hanno sottolineato due aspetti importanti: la necessità di previsioni normative che permettano di attivare strumenti di accertamento standardizzato soltanto quando ci sono indizi certi di reati o di evasione fiscale; la doverosità di strutturare gli strumenti di accertamento presuntivi su elementi il più possibile certi. Considerazioni che si rivelano fondamentali, non solo per comprendere dove bisognerebbe intervenire per adeguare la disciplina tributaria del nostro ordinamento ai migliori standard europei e internazionali, avendo a mente una migliore tutela del cittadino di fronte a poteri di controllo così penetranti, ma anche ai fini della nostra analisi.

Per completare il quadro è utile, allora, riportare un ulteriore elemento, prima di trarre delle conclusioni sulla questione in analisi.

La vera garanzia per il contribuente dall'invasività dei poteri di controllo dell'amministrazione finanziaria attivati da strumenti di accertamento standardizzato - oltre ai principi generali del procedimento amministrativo e tributario, i quali non sono serviti a garantire ancora ragionevolezza e proporzionalità sull'utilizzo degli strumenti di accertamento standardizzato - resta l'obbligatorietà del contraddittorio procedimentale del contribuente prima dell'avviso di accertamento.

Nonostante non ci sia una norma generale sul contraddittorio endoprocedimentale nel procedimento tributario, il d.l. n. 78/2010 ha introdotto l'obbligo del contraddittorio sia per gli studi di settore, con la modifica dell'art. 10, comma 3-bis, della l. n. 146/1998<sup>235</sup>,

---

M. BASILAVECCHIA, *Problemi veri e falsi del nuovo redditometro*, in *Corr. Trib.*, 2013, fasc. 27, pag. 2138 – 2140.

233 A. MARCHESSELLI, *Redditometro e diritti fondamentali: da Garante e giurisprudenza estera un «assist» ai giudici tributari italiani*, cit., fa riferimento alle sentenze della Corte cost. Belgio, 14 febbraio 2013, n. 6 e 14 marzo 2013, n. 39 e alla sentenza CEDU, 14 maggio 2013, 66529/11, «N.K.M. vs Ungheria».

234 A. MARCHESSELLI, *Redditometro e diritti fondamentali: da Garante e giurisprudenza estera un «assist» ai giudici tributari italiani*, cit.

235 Sugli studi di settore per tutti si v. C. NOCERA, *Il contraddittorio nella recente prassi e giurisprudenza in materia di studi di settore*, in P. BORIA (a cura di), *Studi di settore e tutela del contribuente*, Milano, 2010, p. 235 e ss.

sia per il redditometro, con la modifica del comma 7 dell'art. 38 del D.P.R. n. 600/1973<sup>236</sup>.

Il contraddittorio procedimentale, secondo una consolidata giurisprudenza<sup>237</sup>, costituisce, nei procedimenti di accertamento standardizzato, innanzitutto garanzia di plausibilità della rettifica - in quanto permetterebbe il passaggio da una situazione astratta e generale, costruita tramite gli strumenti di accertamento standardizzato, ad una situazione di concreta conoscibilità della capacità contributiva - e lo strumento principale per arginare le forzature che abbiamo ricordato prima e per garantire il contribuente dall'attivazione di poteri istruttori così penetranti<sup>238</sup>.

Il contraddittorio endoprocedimentale, in realtà, non solo consente al contribuente di intervenire nella fase istruttoria quale parte attiva, così da raggiungere una prima tutela nel procedimento, ma è un istituto che permette all'amministrazione finanziaria di reperire tutte le informazioni e i dati concreti sulla capacità impositiva del contribuente, completando così la sua conoscenza della realtà ed avere chiara la situazione di fatto concreta in cui va ad operare<sup>239</sup>.

Il contraddittorio, quindi, è il momento attuativo fondamentale del principio generale del “giusto” procedimento e dell'interesse ordinamentale alla “giusta” determinazione dei tributi, in quanto sia il contribuente, sia l'amministrazione finanziaria concorrono alla corretta qualificazione della materia imponibile<sup>240</sup>.

Le conclusioni che possiamo trarre da questa veloce analisi dell'utilizzo di strumenti di accertamento che elaborano i dati attraverso delle funzioni matematiche e statistiche nei procedimenti tributari sono due.

La prima è che il loro utilizzo certamente migliora l'efficacia e l'efficienza dei controlli fiscali, in quanto sono strumenti capaci di porre all'attenzione dell'amministrazione

---

236 Sugli accertamenti sintetici per tutti si v. D. CONTE, *Aspetti problematici del redditometro e di altri consimili metodi di accertamento come strumenti di contrasto all'evasione*, in *Riv. dir. Trib.*, 2008, I, p. 1138 e ss.; G. RAGUCCI, *Il contraddittorio nei procedimenti tributari*, Torino, 2009, p. 166 e ss..

237 Per tutti si v. Cass., Sez. trib., 9 luglio 2010, n. 16235, con nota di M. DAMIANI, *Nessuna prevalenza tra studi di settore e criterio di tassazione per cassa*, in *Corr. Trib.*, n. 38/2010, p. 3105; Cass., Sez. trib., 22 febbraio 2008, n. 4624, con nota di L.R. CORRADO, *Accertamenti standardizzati e motivazione dell'avviso di accertamento: l'atto è illegittimo in difetto di una adeguata replica alle deduzioni fornite dal contribuente in sede di contraddittorio endoprocedimentale*, in *Dir. Prat. Trib.*, 2008, II, p. 1077.

238 Si v. L.R. CORRADO, *Il contraddittorio endoprocedimentale condiziona l'efficacia probatoria degli studi di settore*, in *Corr. Trib.*, 39, 2015, p. 3972 e ss..

239 Si v. F. GALLO, *Accertamento e garanzie del contribuente*, in *Dir. Prat. Trib.*, 1989, I, p. 67.; R. LUPI, *Metodi induttivi e presunzioni nell'accertamento tributario*, Milano, 1988, p. 331; F. MOSCHETTI, *Avviso di accertamento tributario e garanzie per il cittadino*, in *Dir. Prat. Trib.*, 1983, I, p. 1940.

240 Si v. A. MARCHESELLI, *Accertamenti tributari e difesa del contribuente. Poteri e diritti nelle procedure fiscali.*, cit.; ID., *Il “giusto procedimento” tributario. Principi e discipline*, cit..

finanziaria le situazioni più “rischiose” proprio grazie agli schemi matematici e statistici che sottolineano la regolarità-scostamento dei contribuenti. Una prima considerazione che, nel caso dei *Big data*, assume una valenza e una prospettiva evolutiva importante. Come dicevamo in precedenza le potenzialità conoscitive, anche in termini di “personalizzazione” della rilevazione, e predittive del fenomeno, che andrebbero al di là della semplice analisi descrittiva, non solo ci portano a dire che le funzioni di controllo sarebbero maggiormente efficaci ed efficienti, ma, più in generale, la decisione amministrativa sarebbe ancorata su basi conoscitive certamente maggiori e più solide.

La seconda è che la capacità di questi strumenti di “massificare” il controllo fiscale<sup>241</sup> possono portare l'amministrazione finanziaria ad utilizzare in maniera eccessiva i propri poteri di controllo, con il concreto rischio di sacrificare in maniera non proporzionale le libertà fondamentali dei cittadini. La legislazione tributaria, la giurisprudenza e le indicazioni contenute nel recente provvedimento del Garante della *privacy*, hanno costruito un sistema di tutela per il cittadino retto dal principio di proporzionalità e fondamentalmente endoprocedimentale: la proporzione delle presunzioni statistiche, le quali devono essere ancorate ad elementi certi; l'incapacità dei risultati ottenuti con l'utilizzo di strumenti di accertamento standardizzato di costituire, da soli, una prova giudiziaria, in quanto vengono considerati delle presunzioni semplici; l'obbligo del contraddittorio preventivo con il cittadino contribuente prima dell'avviso di accertamento. Un sistema di tutele che comunque ha dei margini di incertezza legati alla incapacità del cittadino e del giudice di conoscere a fondo i sistemi e i meccanismi matematici e statistici che producono quei risultati.

Certamente un sistema di tutele endoprocedimentale costruito in questa maniera potrebbe reggere anche qualora i *Big data* possano essere utilizzati stabilmente durante l'azione amministrativa anche se con margini di incertezza superiore in merito alla conoscenza diffusa delle tecnologie di elaborazione dei *Big data* e sugli effetti che particolari restrizioni su come e quali dati utilizzare potrebbero esserci sulle potenzialità delle elaborazioni dei *Big data*.

### *3. L'organizzazione amministrativa e i Big data: primi tentativi di organizzare i Big data.*

Dopo avere cercato di comprendere come l'ordinamento italiano ha reagito all'introduzione nel procedimento amministrativo di funzioni matematiche e statistiche,

---

<sup>241</sup> Si v. L. PERRONE, *Il reddito entro verso accertamenti di massa (con obbligo del contraddittorio e dell'invito all'adesione)*, in *Rass. Trib.*, fasc. 4, 2011, p. 887 ss..

al loro utilizzo come presupposto per prendere delle decisioni amministrative e come ha garantito le tutele ai cittadini, con l'obiettivo di cercare di osservare come i *Big data* potrebbero influire sull'azione amministrativa, attraverso un'esperienza concreta che si avvicina per caratteristiche e dinamiche procedurali ad essa, vale la pena riproporre lo stesso schema logico e argomentativo anche per quanto riguarda il rapporto tra *Big data* e organizzazione amministrativa.

Nel corso del primo capitolo abbiamo sottolineato come i *Big data* siano dei dati con delle particolari caratteristiche (volume, velocità e varietà) che, in parte, li differenziano dai dati che tradizionalmente le pubbliche amministrazioni sono abituate a raccogliere, a conservare e a utilizzare<sup>242</sup>. Utilizzare questi dati per qualsiasi tipo di decisione, in particolare per la decisione amministrativa, significa riuscire innanzitutto ad organizzarli e a gestirli. La possibilità che questi possano essere prodotti, archiviati ed elaborati insieme ai dati tradizionali, modifica in parte le caratteristiche del patrimonio informativo delle amministrazioni, rendendolo non più soltanto un insieme di informazioni o di dati, ma anche di metadati, di dati grezzi, utili per produrre nuove informazioni attraverso la loro elaborazione e la loro analisi, ma che non sono immediatamente fruibili.

Abbiamo anche sottolineato come la tendenza all'estrema valorizzazione dei dati, anche se di cattiva qualità, costituisca uno dei cambiamenti culturali maggiormente rilevanti che l'utilizzo dei *Big data* e le loro potenzialità elaborative ha fatto emergere<sup>243</sup>.

Questi due aspetti, quello legato alla natura e alle caratteristiche dei “nuovi” dati e quello essenzialmente culturale, sono certamente i profili problematici più importanti che le amministrazioni pubbliche si trovano a dover affrontare dal punto di vista organizzativo. Il cambiamento culturale verso una maggiore valorizzazione del dato e anche verso una costante datizzazione della realtà ha cominciato a portare le amministrazioni - così come è avvenuto in parte negli esempi riportati nei capitoli precedenti e come vedremo anche poi - a raccogliere tutti i dati possibili su un determinato fenomeno, anche dati di cattiva qualità, con l'obiettivo di sfruttare a pieno le metodologie di elaborazione dei *Big data* per produrre nuove informazioni e nuove conoscenze in modo tale da prendere una decisione il più possibile completa e più aderente alla realtà su cui si vuole intervenire.

La modifica della natura dei dati, la quantità dei dati che è possibile raccogliere e le possibili elaborazioni dei *Big data*, inoltre, obbliga l'amministrazione pubblica a dotarsi di nuove tecnologie per gestire e organizzare i dati. I *Big data*, come abbiamo descritto

---

242 Per tutti diffusamente V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit..

243 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit..

in precedenza, hanno reso inadeguate le tecnologie che supportavano le banche dati tradizionali, cioè quelle relazionali, le quali non sono più in grado di archiviare i dati che hanno le caratteristiche dei *Big data*<sup>244</sup>. L'eterogeneità delle fonti e delle forme, la loro enorme quantità, la crescita esponenziale della loro produzione costringono le amministrazioni a ricorrere a nuovi strumenti tecnologici che siano in grado di archiviare e produrre i *Big data* senza eccessiva difficoltà. La possibilità di utilizzare pienamente i *Big data*, quindi, passa dalla capacità delle amministrazioni pubbliche di sapere rinnovare i loro strumenti tecnologici di conservazione e di elaborazione, adeguandoli alle tecnologie dei *Big data*.

Allora è utile analizzare quali sono state le prime esperienze di organizzazione e gestione dei *Big data* nelle amministrazioni pubbliche in maniera tale da potere osservare quali sono, tra gli ambiti di influenza che abbiamo descritto, quelli che più di tutti si sono mostrati come ambiti problematici per le amministrazioni, con lo scopo di costruire una base concreta di esperienze da cui sia possibile fare emergere il rilievo giuridico all'interno del rapporto tra *Big data* e organizzazione amministrativa.

### *3.1. Le esperienze di organizzazione dei Big data: ISTAT, INPS e INAIL.*

Dalla ricostruzione fatta nel paragrafo precedente sono emersi due grossi ambiti problematici che le amministrazioni devono affrontare, dal punto di vista dell'organizzazione dei *Big data*: il bisogno di nuove tecnologie di raccolta, conservazione e elaborazione dei dati e la necessità di nuovi metodi di organizzazione dei dati per poterne utilizzare le potenzialità e per poterli utilizzare in modo del tutto nuovo.

Le amministrazioni pubbliche che, per prime, si sono trovate a dovere affrontare le problematiche che abbiamo descritto nel paragrafo precedente sono quelle che più di tutte si trovano a possedere e a gestire un ingente quantità di dati e di informazioni: stiamo parlando dell'Istituto nazionale di statistica (ISTAT), dell'Istituto nazionale della previdenza sociale (INPS) e dell'Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro (INAIL).

I tre Istituti nazionali sono chiamati, infatti, dal nostro ordinamento a svolgere funzioni fondamentali e a gestire, ognuno per la propria finalità istituzionale, moltissimi dati.

---

244 A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, Maggioli, Bologna, 2013, cit..

L'ISTAT, ma tutto il Sistema statistico nazionale (SISTAN), si trova nella posizione di frontiera nell'organizzazione dei *Big data*, e le finalità istituzionali dell'Istituto, delineati all'art. 15 del d.lgs. n. 322/1989, sono quelle che più di tutte si trovano a subire gli effetti del fenomeno dei *Big data*<sup>245</sup>. La legge affida all'ISTAT principalmente compiti inerenti l'esecuzione dei censimenti, delle rilevazioni statistiche più importanti sui fenomeni economico-sociali del Paese e la predisposizioni e la diffusione di studi e ricerche basate su queste rilevazioni. Una funzione di produzione continua di conoscenza pubblica che viene affiancata a compiti di indirizzo, di coordinamento, di valutazione dell'adeguatezza dell'attività statistica e di assistenza tecnica per tutti gli uffici e gli enti che fanno parte del Sistema statistico nazionale.

Un'attività conoscitiva che non può non essere investita fortemente dai nuovi problemi di gestione dei dati e dalle nuove domande che i nuovi fenomeni conoscitivi pongono: che fare dei *Big data*; come raccogliere e conservare i dati, come valorizzare le nuove possibilità di comprensione della realtà e con quali finalità conoscitive.

La legge, inoltre, affida all'ISTAT anche compiti molto più tecnici: la predisposizione delle nomenclature e delle metodologie di base per la classificazione e la rilevazione dei fenomeni di carattere demografico, economico e sociale, le quali sono vincolanti per gli enti ed organismi facenti parte del Sistema statistico nazionale e la promozione e lo sviluppo informatico a fini statistici degli archivi gestionali e delle raccolte di dati amministrativi. Compiti tecnici che obbligano l'Istituto a confrontarsi con il fenomeno dei *Big data* anche dal punto di vista tecnologico e scientifico, in particolare in merito alla necessità di dovere rinnovare le proprie tecnologie e di confrontarsi con i metodi di elaborazione dei *Big data* che utilizzano correlazioni non lineari e che hanno la tendenza ad abbandonare le analisi basate sul rapporto causa-effetto.

L'ISTAT, quindi, si trova in una posizione di frontiera su questo tema, rispetto alle altre amministrazioni pubbliche e ha, in potenza, anche una certa capacità di poter direzionare ricerca e metodologie statistiche verso una strutturazione più adatta ad un maggiore utilizzo dei *Big data*, ma non è l'unica amministrazione pubblica ad esserlo.

Nella stessa posizione di frontiera si trova l'Istituto nazionale della previdenza sociale (INPS), l'ente pubblico non economico la cui attività principale consiste nella liquidazione e nel pagamento delle pensioni e delle indennità di natura previdenziale e

---

245 Sul ruolo delle garanzie organizzative e professionali ai fini della qualità dell'informazione statistica si v. M. P. GUERRA, *Statistica pubblica e funzione conoscitiva*, in F. MERLONI (a cura di), *L'informazione delle pubbliche amministrazioni*, Rimini, 2002; F. GIGLIONI, *Le soluzioni istituzionali alla qualità dell'informazione statistica*, in *Dir. Pubbl.*, 2005. Sulla statistica delle pubbliche amministrazioni: L.L. RONDINI, *Storia della statistica pubblica in Italia*, Milano, 2003; C. GRASSETTI, *Statistica per la pubblica amministrazione*, Padova, 2008.

di natura assistenziale a cui si aggiungono anche attività di vigilanza, con i relativi poteri di controllo<sup>246</sup>.

L'INPS, secondo le informazioni ufficiali<sup>247</sup>, gestisce la previdenza e l'assistenza sociale di oltre 40,7 milioni di utenti, di cui 23,4 milioni di lavoratori (l'82% della popolazione occupata in Italia), 1,4 milioni di imprese, 16 milioni di pensionati; 4,4 milioni di persone che ricevono prestazioni a sostegno del reddito. Un patrimonio informativo di 25 terabyte di dati. Un Istituto, quindi, in possesso di una delle banche dati più ingenti del nostro Paese e che è costretto a confrontarsi con le potenzialità del fenomeno dei *Big data*: si pensi a come potrebbero cambiare le attività di controllo dell'Istituto o come potrebbero migliorare i servizi offerti ai cittadini o, ancora, a come l'elaborazione dei *Big data* potrebbe essere utile a migliorare ed estendere la tutela previdenziale e assistenziale.

Nella stessa situazione dell'INPS si trova anche l'Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro (INAIL), che gestisce l'assicurazione obbligatoria contro gli infortuni sul lavoro e le malattie professionali.

L'INAIL svolge attività di natura eterogenea che vanno dallo studio delle situazioni di rischio, agli interventi di prevenzione nei luoghi di lavoro, alle prestazioni sanitarie ed economiche, alla cura, riabilitazione e reinserimento nella vita sociale e lavorativa nei confronti di coloro che hanno subito danni fisici a seguito di infortunio o malattia professionale<sup>248</sup>. Anche in questo caso la comparsa dei *Big data* hanno portato l'ente a doversi confrontare con problematiche e necessità nuove.

È utile fare due precisazioni. Innanzitutto occorre sottolineare che quando parliamo di esperienze concrete di gestione di *Big data*, parliamo di progetti ancora iniziali o ancora in corso che terremo conto nella misura in cui ci è utile per delineare quali sono gli interventi generali che accomunano i primi interventi concreti di organizzazione dei *Big data*. In secondo luogo, che facciamo riferimento soltanto alle esperienze compiute dagli Istituti nazionali che abbiamo elencato in quanto sono quelle più strutturate, ma

---

246 Per un approfondimento sull'INPS si v. S. CIANELLA – L. ZOPPOLI, *Istituto Nazionale di Previdenza Sociale*, in *D. disc. Priv., sez. comm.*, VIII, Torino, Utet, 1992, p. 9; P. BOER, *L'attività di accertamento, vigilanza ed ispezione degli Enti previdenziali*, in *Mass. Giur. Lav.*, 1996, p. 679. Per una panoramica generale sulla previdenza sociale si v. A. LA SPINA, *La protezione sociale*, in S. CASSESE (a cura di), *Trattato di diritto amministrativo*, p. 823 e ss.; M. CINELLI, *Sicurezza sociale*, in *Enc. Dir.*, XLII, Milano, 1990, p.4; ID., *Diritto alla previdenza sociale*, Torino, 1998; S. GIUBBONI – M. CINELLI, *Il diritto della sicurezza sociale in trasformazione*, Torino, 2005.

247 Per i dati si è consultato il sito dell'Istituto nazionale di previdenza sociale (INPS): [https://www.inps.it/portale/default.aspx?iMenu=11&bi=1\\_1&link=L%27ISTITUTO](https://www.inps.it/portale/default.aspx?iMenu=11&bi=1_1&link=L%27ISTITUTO)

248 Sull'INAIL si v. P. CIANELLA – S. CIANELLA, *Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro*, in *D. disc. Priv., sez. comm.*, VIII, Torino, 1992, p. 31; P. BOER, *L'attività di accertamento, vigilanza ed ispezione degli Enti previdenziali*, cit.. Sulla disciplina degli infortuni sul lavoro e delle malattie professionali si v. G. FERRARI – G. FERRARI, *Infortuni sul lavoro e malattie professionali*, Padova, IV, 2002.

non mancheremo di evidenziare anche le esperienze e le decisioni in merito di altre amministrazioni.

Dalla comparazione e dalle analisi delle prime esperienze dell'ISTAT, dell'INPS e dell'INAIL è possibile individuare tre grandi filoni di intervento che hanno accomunato, anche se in maniera più o meno intensa, tutte le esperienze in analisi: l'ammodernamento tecnologico con il fine di riuscire a conservare ed elaborare i *Big data*; la tendenza verso politiche di integrazione e fusione delle banche dati esistenti; la volontà di introdurre, in via ancora sperimentale, i *Big data* nelle ricerche, nelle analisi e a sostegno del processo decisionale.

Politiche che mettono in evidenza quali sono le necessità di ammodernamento e adattamento al fenomeno più urgenti, che le amministrazioni hanno percepito, e quali sono i primi rilievi giuridici in termini di organizzazione dei dati.

### *3.1.1. Le politiche di organizzazione dei Big data dell'Istituto nazionale di Statistica (ISTAT).*

L'ISTAT, tra gli Istituti nazionali in analisi, è stato certamente quello più attivo e più interessato, per le proprie finalità istituzionali e per la vicinanza scientifica al fenomeno, ad affrontare le problematiche di organizzazione e sfruttamento dei dati che i *Big data* hanno sollevato.

L'ISTAT ha portato avanti, innanzitutto, importanti politiche di ammodernamento tecnologico delle proprie strumentazioni di conservazione e di analisi dei dati. Non solo ha recentemente migrato il suo patrimonio di dati verso un *database machine Oracle Exadata* e ha utilizzato tecnologie *PowerPivot* per le attività di *Business Intelligence* durante l'ultimo censimento della popolazione<sup>249</sup> – tecnologie che daranno la possibilità all'Istituto di conservare ed elaborare i *Big data* in maniera adeguata - ma ha intenzione di dotarsi di una strumentazione informatica ancora più all'avanguardia, in quanto ha avviato un accordo con l'Università degli studi di Pisa per lo studio di nuove tecniche informatiche adatte allo scopo, con l'obiettivo di migliorare ancora le proprie capacità elaborative<sup>250</sup>.

---

249 Per un approfondimento sulle tecnologie di conservazione e di elaborazione dei *Big data* utilizzate dall'ISTAT si v. diffusamente il *Programma statistico nazionale (PSN) 2014-2016 – Aggiornamento 2016*.

250 La notizia è stata riportata nelle news del sito ISTAT. Si può facilmente consultare l'intervista in cui sono state riportate queste informazioni sul sito [http://www.istat.it/it/files/2013/02/Big\\_Data.pdf](http://www.istat.it/it/files/2013/02/Big_Data.pdf).

La più rilevante politica di organizzazione dei *Big data* compiuta dall'ISTAT e che merita una particolare attenzione, però, è quella riguardante la fusione e l'integrazione delle banche dati a propria disposizione. L'obiettivo dell'Istituto è quello di incrementare la quantità di dati a propria disposizione e avvicinarne la quantità ad una dimensione tale da poterne sfruttare le potenzialità di elaborazione.

Emblematiche sono state, infatti, le scelte che l'ISTAT ha fatto, in maniera progressiva, negli ultimi Programmi statistici nazionali (dal PSN 2011-2013) e che sono ben descritte, e ulteriormente arricchite, all'interno del Programma statistico nazionale (PSN) 2014-2016 – Aggiornamento 2016<sup>251</sup>.

Il PSN vorrebbe ampliare ulteriormente il Sistema Integrato dei Microdati (SIM)<sup>252</sup>, un sistema informativo all'interno del quale l'ISTAT, nel corso degli anni, ha fatto confluire circa 60 archivi di fonte amministrativa, per circa 500 milioni di record all'anno<sup>253</sup>. Con il nuovo PSN, l'Istituto di statistica ha intenzione di estendere la grandezza del SIM e di includere al suo interno ulteriori banche dati, moltiplicando enormemente la capacità dell'Istituto di elaborare i dati attingendo da tutte le fonti disponibili del sistema amministrativo<sup>254</sup>. Un obiettivo che darebbe all'ISTAT un forte potere informativo e una migliore capacità elaborativa.

Prima di continuare la nostra analisi è utile, però, riportare alcune considerazioni che il Garante della *privacy* ha fatto in merito a questo punto del Piano statistico nazionale: un parere obbligatorio in quanto il PSN, dopo essere stato deliberato dal Comitato di indirizzo e coordinamento dell'informazione statistica nazionale (COMSTAT), deve essere sottoposto ai pareri della Commissione per la garanzia della qualità dell'informazione statistica, della Conferenza unificata Stato-Regioni-Autonomie locali e del Garante per la protezione dei dati personali.

---

251 Il Programma statistico nazionale (PSN) è l'atto normativo che, in base all'art. 13 del d.lgs. n. 322/1989 e successive integrazioni, stabilisce le rilevazioni statistiche di interesse pubblico affidate al Sistema statistico nazionale e i relativi obiettivi informativi.

252 Il Sistema Integrato dei Microdati (SIM) è stato inserito per la prima volta nel PSN 2011-2013 come Registro statistico nazionale degli individui. Il SIM non è realizzato per produrre un risultato statistico autonomo ma viene utilizzato dall'ISTAT quale archivio intermedio per la realizzazione di numerosi lavori statistici.

253 Per queste cifre si v. ISTAT, *Linee strategiche del censimento permanente della popolazione e delle abitazioni, metodi, tecniche e organizzazione*, 2014.

254 In particolare, il Casellario dei pensionati, la Base dati anagrafica dell'Anagrafe Tributaria, la Banca dati lavoratori interinali e parasubordinati, l'Archivio anagrafico dei lavoratori extracomunitari, l'Anagrafe nazionale degli studenti Universitari, l'Anagrafe degli studenti, la banca dati Contratti di locazione, la banca dati sul Parco autoveicoli circolanti, l'Archivio dei numeri di telefonia fissa privati. L'ISTAT vorrebbe aggiungere, tra il resto, anche l'Anagrafe Nazionale delle Persone Residenti, la banca dati sulle Tessere Sanitarie, l'Archivio delle certificazioni telematiche delle malattie, l'Archivio degli esiti degli esami di passaggio dei cicli scolastici, l'Anagrafe virtuale statistica, la Base integrata di microdati amministrativi e statistici per l'analisi dell'occupazione.

Con il Parere n. 566 del 29 ottobre 2015, il Garante ha sottolineato, nonostante il parere sia stato sostanzialmente positivo, come l'integrazione delle banche dati con quelle esistenti nel SIM, che l'ISTAT intende portare avanti, possa duplicare e accentrare la quasi totalità delle banche dati pubbliche presso un solo soggetto e spesso non sono direttamente finalizzati ad un determinato risultato statistico<sup>255</sup>. Il Sistema Integrato dei Microdati, infatti, non è realizzato per produrre un risultato statistico autonomo ma viene utilizzato dall'ISTAT quale archivio intermedio per la realizzazione di numerosi lavori statistici, tra cui il censimento permanente. Il Garante teme, infatti, da un lato, che ci possano essere rischi di profilazione di massa sull'intera popolazione, in quanto il SIM permette di selezionare campioni ragionati, caratterizzati da unità campionarie già ampiamente corredate di informazioni che ricostruiscono il profilo e perfino la personalità dei cittadini, dall'altro, che vengano lesi il principio di finalità, appunto perché la raccolta dei dati personali non sono legati ad una particolare analisi statistica, e il principio di necessità, in quanto i dati raccolti non vengono cancellati dopo il loro utilizzo, in quanto questo non è chiaro. A seguito di queste considerazioni il Garante si è innanzitutto riservato di valutare con attenzione i rischi specifici di questi trattamenti e non si accontenta di una valutazione documentale, ma specifica che si tratterà di fare delle valutazioni sulla base di elementi «da acquisire nell'ambito di specifici approfondimenti istruttori, anche di carattere ispettivo»<sup>256</sup>.

In secondo luogo, il Garante afferma che l'utilizzo dell'integrazione delle banche dati a fini statistici, considerando che i dati erano stati raccolti per finalità differenti (finalità amministrative), richiedono, infatti, l'introduzione di uno specifico quadro di regole a tutela degli interessati, ulteriori rispetto a quanto allo stato previsto dall'ISTAT. Ciò, al fine di assicurare la correttezza e la liceità del trattamento posto in essere e di individuare idonee garanzie in relazione alla natura e alla qualità dei dati, alle modalità del trattamento e agli effetti che può determinare sugli interessati, nonché alle misure di sicurezza. Il Garante, quindi, ha concluso che l'ISTAT debba sottoporre ad una sua verifica preliminare, ai sensi dell'art. 17 del Codice, i trattamenti di dati personali previsti dall'integrazione del SIM e che tali trattamenti potranno essere avviati solo all'esito dell'apposita verifica preliminare.

Oltre all'integrazione delle banche dati, l'ISTAT sta cominciando anche ad utilizzare i *Big data* nelle proprie ricerche e nelle proprie analisi statistiche e, nei casi più sperimentali, ad elaborare i *Big data*.

---

255 Garante per la protezione dei dati personali, *Parere sullo schema di Programma statistico nazionale (Psn) 2014-2016 - Aggiornamento 2016*, del 29 ottobre 2015, n. 566.

256 Garante per la protezione dei dati personali, *Parere sullo schema di Programma statistico nazionale (Psn) 2014-2016 - Aggiornamento 2016*, del 29 ottobre 2015, n. 566.

È sempre più frequente, infatti, che i *Big data* vengano utilizzati come tecniche alternative per la raccolta dei dati, come nel caso in cui è stato utilizzato internet per le rilevazioni sull'utilizzo delle tecnologie ICT da parte di imprese e pubbliche amministrazioni o per la costruzione dell'Indice dei Prezzi al Consumo<sup>257</sup>.

Molto più interessante anche se solo in via sperimentale è l'utilizzo integrato dei *Big data* con i dati di indagine statistica tradizionale. Vale la pena di riportare un esempio emblematici di questo nuovo e ancora poco indagato approccio dell'Istituto e riguarda l'utilizzo delle interrogazioni sui motori di ricerca per fare previsioni e proiezioni sul mercato del lavoro e per costruire degli Indicatori del Mercato del Lavoro. L'esperimento statistico è stato quello di valutare la possibilità di utilizzare, come informazioni ausiliarie, le serie storiche delle interrogazioni relative a termini come "lavoro", "offerte di lavoro" e simili, ottenute da *Google Trends*, al fine per produrre sia delle stime anticipate e delle previsioni del di tasso mensile di disoccupazione le previsioni, sia delle stime per piccole aree dello stesso indicatore<sup>258</sup>.

L'ISTAT, infine, con il Programma statistico nazionale (PSN) 2014-2016 – Aggiornamento 2016, sta continuando un esperimento, avviato con l'Aggiornamento 2015 al PSN in analisi, per utilizzare nelle analisi e nelle ricerche statistiche soltanto i *Big data* in sostituzione dei dati di indagine tradizionali. Si tratta del primo vero esperimento di elaborazione dei *Big data* come lo abbiamo descritto e analizzato in questo lavoro e merita particolare attenzione, anche perché è stato anch'esso oggetto di attenzione da parte del Garante della *privacy*.

L'esperimento riguarda l'uso dei dati di telefonia mobile per stimare i flussi di popolazione inter-comunale. Con il progetto *Persons and Places*, infatti, l'ISTAT utilizzerà il c.d. “*call detail record*”, cioè il numero progressivo che ogni compagnia telefonica assegna all'utente nel momento in cui effettua una chiamata e che utilizza l'ora e l'area da dove è stata effettuata la chiamata per stimare, a livello aggregato, i flussi di mobilità delle persone. L'indagine dell'ISTAT segmenterà la popolazione in quattro categorie analizzando i *Big Data* provenienti dai cellulari: residenti stanziali, temporaneamente dimoranti, pendolari giornalieri e visitatori occasionali. I flussi sono utilissimi per gestire in maniera ottimale alcuni servizi, come ad esempio il trasporto

---

257 Per informazioni ulteriori si G. ALLEVA, *Conservazione e accessibilità dei microdati prodotti dall'ISTAT e cenni all'ambito del SISTAN*, Audizione del Presidente dell'ISTAT presso la Commissione per la garanzia della qualità dell'informazione statistica, 12 febbraio 2016; D. DONVITO, *Big data e Statistica Ufficiale: le strategie ISTAT*, in <http://docplayer.it/958327-Big-data-e-statistica-ufficiale-la-strategia-istat.html>.

258 G. ALLEVA, *Conservazione e accessibilità dei microdati prodotti dall'ISTAT e cenni all'ambito del SISTAN*, cit.; D. DONVITO, *Big data e Statistica Ufficiale: le strategie ISTAT*, in <http://docplayer.it/958327-Big-data-e-statistica-ufficiale-la-strategia-istat.html>.

pubblico locale, e programmare opportune misure d'intervento della Protezione Civile in caso di necessità.

Il Garante, riprendendo le considerazioni del parere precedente<sup>259</sup>, ha sottolineato che si riserverà di verificare la corretta utilizzazione dei dati di telefonia, in particolare di verificare se questi dati restino anonimi, così come è stato assicurato dall'ISTAT. Il rischio, continua il Garante, è quello che grazie alle nuove tecnologie e alle nuove tecniche di analisi, di elaborazione e di interconnessione dei dati risulta possibile "re-identificare" un interessato anche attraverso informazioni apparentemente anonime.

Le considerazioni del Garante sull'integrazione delle banche dati e sull'elaborazione dei dati personali, anche anonimi, fanno emergere, da un lato, come il rapporto *Big data* e *privacy* resti il rilievo giuridico più ricorrente nell'analisi che stiamo compiendo sui *Big data*, dall'altro, come sia complicato nel nostro ordinamento trovare l'equilibrio tra introdurre questi strumenti di elaborazione dei dati e garanzie per il cittadino, senza atrofizzare le potenzialità dell'uno o reprimere eccessivamente le garanzie per l'altro.

Considerare l'ISTAT come una delle amministrazioni di frontiera sul tema è una valutazione che viene anche supportata dal costante impegno dell'Istituto a confrontarsi con il fenomeno dei *Big data*. L'ISTAT ha deciso, infatti, di costituire un gruppo di esperti che studi il fenomeno e il suo impatto sulla statistica ufficiale. La Commissione ha due obiettivi principali. Da una parte, si propone di definire "cosa" utilizzare, in termini di dati, metodi e tecnologie afferenti al mondo *Big Data*, nel processo di produzione dell'informazione statistica attuale. Rientra in tale obiettivo, la valutazione dell'utilizzo delle sorgenti di *Big Data* come sorgenti ausiliarie che integrano le sorgenti tradizionali. Dall'altra parte, vuole definire "come", eventualmente, cambiare il processo di produzione dell'informazione statistica. A tal fine, le sorgenti di *Big Data* saranno valutate sia come sorgenti aggiuntive a quelle tradizionali per studiare fenomeni allo stato non considerati nell'ambito della produzione statistica ufficiale, sia come sorgenti sostitutive a quelle tradizionali, dove ne sia provata una maggiore efficacia dal punto di vista dei costi senza una sostanziale perdita di qualità dell'informazione prodotta<sup>260</sup>.

In realtà la necessità di indagare il rapporto tra *Big data* e statistica ufficiale è una necessità di rilievo internazionale. L'High-level group for the modernisation of statistical production and services (HLG), gruppo internazionale costituito da presidenti di organizzazioni statistiche, di cui fa parte anche l'ISTAT, il quale è la sede in cui, a livello internazionale, vengono decise e indirizzate le strategie di modernizzazione della

---

259 Garante per la protezione dei dati personali, *Parere sul PSN 2014-2016 aggiornamento 2015-2016*, 18 settembre 2014 n. 411.

260 È possibile reperire questa informazione on-line in M. SCAPPIERO, *Esperti a confronto per orientare le scelte dell'ISTAT sui Big data*, in [http://www.istat.it/it/files/2013/06/big\\_data.pdf](http://www.istat.it/it/files/2013/06/big_data.pdf).

produzione statistica del mondo, ha rilevato l'importanza del fenomeno dei *Big Data* e la necessità per le organizzazioni statistiche di rivedere le loro funzioni alla luce del fenomeno stesso. A tale scopo l'HLG sta sponsorizzando una serie di progetti di collaborazione internazionale per capire meglio come sfruttare le potenzialità dei *Big Data* nella produzione di statistiche ufficiali<sup>261</sup>.

*3.1.2. Le politiche di organizzazione dei Big data dell'Istituto nazionale di previdenza sociale (INPS) e dell'Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro (INAIL);*

Dopo avere elencato quali sono state le politiche di organizzazione dei *Big data* portate avanti dall'ISTAT, è utile allora elencare quelle portate avanti, anche se in maniera meno strutturata rispetto alla prima, dall'Istituto nazionale di previdenza sociale (INPS) e dall'Istituto nazionale per l'assicurazione contro gli infortuni sul lavoro (INAIL).

Le politiche di modernizzazione tecnologica dell'INAIL sembrano aver avuto come obiettivo principale la razionalizzazione della tecnologia esistente, piuttosto che un radicale rinnovamento tecnologico, utile per elaborare i *Big data*<sup>262</sup>. L'Istituto, in vista della volontà di integrare i dati inerenti il *welfare* nazionale, come vedremo poi, sta lavorando ad una piattaforma su cui implementare la banca dati unitaria<sup>263</sup>. Una scelta che cerca di mettere insieme una mole di dati tale da poter raggiungere delle potenzialità di elaborazione degli stessi elevata, simile all'elaborazione dei *Big data*, piuttosto che una che cerca di valorizzare i risultati delle analisi degli stessi.

L'INPS, invece, è da tempo in possesso delle tecnologie capaci di gestire un gran numero di dati, in quanto l'INPS, come abbiamo detto nei paragrafi precedenti, possiede il sistema informativo più grande del Paese. L'Istituto, infatti, si è dotato già dal 2000 di un sistema di *data warehouse* che, alimentato dai sistemi di gestione, genera in automatico liste, *report* e KPI con finalità operative, strategiche e finanziarie. Come abbiamo potuto vedere nel primo capitolo le tecnologie che permettono il *data warehouse* sono certamente adatte a produrre analisi ed elaborazioni su una mole di dati molto elevata, ma restano ancora poco utili per analisi predittive su basi di dati molto eterogenei<sup>264</sup>.

---

261 Per informazioni più dettagliate sui progetti che l'HLG sta portando avanti nel 2015 e su quelli compiuti nel 2014 si v. <http://www1.unece.org/stat/platform/display/bigdata/2015+project>.

262 Per approfondire si v. C. DE PADOVA, *Big data, la nuova frontiera dell'IT*, in *Riv. Inf. Mal. Prof.*, 2013, fasc. 1-2, p. 247.

263 Ibidem.

264 G. DEL MONDO, *Big Data e Analytics: modelli, analisi e previsioni, per valorizzare l'enorme patrimonio informativo pubblico*, in

Se dal punto di vista strettamente tecnologico le politiche dei due Istituti in analisi non sono poi state così attente alle nuove necessità tecnologiche fatte emergere dai *Big data*, in merito alle politiche di integrazione delle banche dati e a quelle in merito ai primi esperimenti di utilizzo dei *Big data*, i due Istituti, così come l'ISTAT, hanno compiuto delle scelte interessanti.

L'INAIL, al fine di monitorare l'efficacia, l'attuazione e i risultati delle proprie politiche assicurative e al fine di prepararsi anche per gestire i *Big data*, ha deciso di integrare tutti i dati a propria disposizione per realizzare una banca dati unica che costituisca la base di partenza del Welfare Information Capital Center (WICC), il quale avrà l'obiettivo di censire, qualificare e integrare tutte le fonti disponibili per lo scopo. L'Istituto, infatti, ha intenzione di integrare i dati in merito ai lavoratori e alle aziende, che detiene copiosamente, insieme ai dati inerenti aspetti socio-ambientali con l'obiettivo di effettuare delle analisi che sfruttino le elaborazioni e le correlazioni statistiche dei *Big data*. Considerando le finalità istituzionali dell'Istituto, l'elaborazione dei *Big data* potrebbe migliorare la ricostruzione da parte dell'INAIL del legame che c'è tra infortuni/malattie professionali e altri fattori (contesto locale, comportamenti aziendali, generatori di stress, stile di vita del lavoratore, azioni di prevenzione etc.), al fine di migliorare le decisioni di tutela e la capacità dell'Istituto di tutelare il lavoratore<sup>265</sup>.

L'INPS, invece, innanzitutto utilizza le analisi prodotte con il *data warehouse* per migliorare i processi decisionali dei suoi organi politici, in quanto esse consentono di descrivere fenomeni sociali, attuali e futuri, permettendo agli organi decisionali di avere a disposizione strumenti di monitoraggio adeguati e, conseguentemente, di adottare azioni correttive a livello organizzativo, procedurale e amministrativo. Un utilizzo dei dati che si è rivelato fruttuoso soprattutto per le finalità di vigilanza e di controllo<sup>266</sup>.

Oltre alle analisi tradizionali che l'INPS compie sulla grande quantità di dati che essa dispone, l'Istituto che si occupa di previdenza sociale ha compiuto recentemente un importantissimo esperimento di elaborazione e di analisi previsionale con i *Big data*. L'INPS, in collaborazione con il Gruppo Daman, è riuscita a stimare in anticipo il ricorso alla cassa integrazione da parte di 356.057 aziende assicurate, utilizzando il software Kxen e con un buon margine di correttezza. Il test era sull'ultimo trimestre del 2012 e ha previsto il ricorso alla Cassa Integrazione Guadagni da parte di 24.236

---

[https://www.inps.it/docallegati/InpsComunica/ForumPA2013/Documents/Del%20Mondo-ForumPA\\_2013\\_definitiva.pdf](https://www.inps.it/docallegati/InpsComunica/ForumPA2013/Documents/Del%20Mondo-ForumPA_2013_definitiva.pdf); Sulle tecnologie di data warehouse e sulla loro incapacità di elaborare nel migliore dei modi i *Big data* si v. A. REZZANI, *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, Maggioli, Bologna, 2013, cit..

265 C. DE PADOVA, *Big data, la nuova frontiera dell'IT*, in *Riv. Inf. Mal. Prof.*, 2013, fasc. 1-2, p. 253.

266 G. DEL MONDO, *Big Data e Analytics: modelli, analisi e previsioni, per valorizzare l'enorme patrimonio informativo pubblico*, cit..

aziende. La realtà osservata a posteriori ha dimostrato che 334.408 aziende non hanno fatto alcuna domanda, mentre 21.667 lo hanno fatto. Il risultato della previsione è dunque risultato esatto al 93,1 per cento<sup>267</sup>.

L'esperimento dimostra come il passo avanti che si può compiere con modelli predittivi è enorme. L'INPS non solo potrebbe migliorare le sue previsioni sulla CIG, in quanto un suo aumento sarebbe capace di far variare di qualche miliardo il bilancio dell'Istituto, ma anche le sue previsioni sulle pensioni che l'INPS si trova a fare e che, con l'ISTAT, fornisce direttamente dati ed elaborazioni a EUROSTAT, facendo parte del Sistema statistico nazionale (SISTAN) ed essendo una delle 19 autorità statistiche nazionali.

### *3.2. L'estrema valorizzazione del patrimonio informativo come fil rouge delle politiche di organizzazione dei Big data.*

Dalla ricostruzione che abbiamo appena fatto in questi paragrafi possiamo affermare che le esperienze concrete di primo utilizzo dei *Big data* e le politiche infrastrutturali (tecnologiche e di integrazione delle banche dati) degli Istituti nazionali che abbiamo voluto riportare, in quanto esempi concreti di come il fenomeno dei *Big data* sta cambiando il modo in cui le amministrazioni organizzano e gestiscono i propri dati, non sono gli unici momenti in cui l'amministrazione italiana si sia affacciata al tema del rapporto tra *Big data* e organizzazione amministrativa, intesa qui come organizzazione dei dati pubblici e privati in proprio possesso.

Oltre all'ISTAT, anche altre amministrazioni si stanno dotando di Commissioni di esperti che possano innanzitutto mappare e comprendere il fenomeno e, in secondo luogo, fornire delle soluzioni concrete per utilizzare i *Big data* secondo le finalità istituzionali dell'amministrazione protagonista dell'iniziativa.

Vale la pena riportare qui la recente decisione del Ministero dell'Istruzione, dell'Università e della Ricerca di costituire un Gruppo di lavoro sui *Big data*. Con il decreto ministeriale del 27 gennaio 2016 n. 34, il Ministero nell'ambito dell'attuazione del Piano nazionale per la scuola digitale<sup>268</sup>, ha affidato a questo gruppo il compito di mappare, in chiave comparata anche a livello europeo e internazionale, i principali centri universitari e non che si occupano di utilizzo dei *Big data* nell'ambito della

---

267 Un esperimento riportato da Davide Colombo sul Sole 24 ore del 17 marzo 2013 e consultabile online su <http://24o.it/5ocJ9>

268 Secondo l'art. 1, comma 56 della legge n. 107/2015 (c.d. "La buona scuola") prevede che il Ministero dell'Istruzione, dell'Università e della Ricerca adotti un Piano nazionale per la scuola digitale in cui siano riportate anche le strategie per garantire stabilmente l'accesso e la riutilizzabilità dei dati pubblici del sistema nazionale di istruzione e formazione in formato aperto.

formazione e nella costruzione di competenze; di identificare le possibili misure che il sistema formativo italiano potrebbe adottare per adeguarsi alle migliori prassi europee ed internazionali valorizzando le proprie specificità; effettuare la ricognizione delle banche dati del Ministero, relative al settore dell'istruzione scolastica, dell'Università, della ricerca scientifica e dell'alta formazione artistica e musicale e proporre delle azioni che il Ministero potrebbe intraprendere al fine di valorizzare il proprio patrimonio informativo e rendere possibile una loro valorizzazione per migliorare le scelte di *policy*.

La valorizzazione del patrimonio informativo pubblico e il suo utilizzo per migliorare la consapevolezza delle decisioni politiche e amministrative si conferma, anche alla luce degli obiettivi che il MIUR si è prefissato attraverso la costituzione di questo gruppo di esperti, come il perno centrale delle *policy* in materia di organizzazione dei dati, anche in un'ottica dei *Big data*. La scelta del MIUR va contestualizzata, infatti, nella più generale politica di valorizzazione del patrimonio informativo pubblico inserita all'interno dell'Agenda digitale italiana e guidata dalla Agenzia per l'Italia Digitale (AgID). La Strategia per la crescita digitale 2014-2020, infatti, va in questa direzione e tutte le politiche che vengono considerate prioritarie tentano di mettere le amministrazioni pubbliche nelle condizioni infrastrutturali e culturali per organizzare in maniera differente e più efficace i dati pubblici e privati a loro disposizione<sup>269</sup>. Bisogna sottolineare, però, come le politiche di valorizzazione, da un lato, sono la condizione necessaria (ma non sufficiente) per rendere il sistema amministrativo italiano capace di affrontare le sfide tecniche, normative e culturali che il fenomeno dei *Big data* sta producendo, dall'altro, non sono perfettamente tarate sul fenomeno in analisi. La Strategia dell'AgID, infatti, tende più a razionalizzare e organizzare l'esistente che a prospettare una diversa valorizzazione e gestione dei dati, più funzionale alle caratteristiche del fenomeno e, forse, è una tendenza anche fisiologica. Interessante, però, si sembra sottolineare come il legame tra valorizzazione del patrimonio informativo – *Big data* – migliore decisione politica e amministrativa sia molto presente nella Strategia quando si parla di *Smart cities*. Nella Strategia per la crescita digitale, il contesto urbano viene delineato come il contesto perfetto all'intero del quale valorizzare i dati, cioè il contesto in cui è possibile raccogliere, conservare e utilizzare i dati così

---

<sup>269</sup> Le politiche più importanti della Strategia fanno riferimento a moltissimi settori dell'amministrazione italiana. Innanzitutto tra le politiche infrastrutturali che ci interessano maggiormente, le più importanti sono quelle riferibili al Sistema pubblico di connettività, al Servizio Pubblico d'Identità Digitale (SPID), al consolidamento di *datacenter* e *cloud computing*, all'Anagrafe Popolazione Residente agli Open data. In secondo luogo interessanti sono quelle che mirano ad una generalizzata digitalizzazione di tutti i settori più importanti: dalla scuola, alla sanità e all'agricoltura. Per approfondire le politiche si v. Presidenza del Consiglio dei ministri, *Strategia per la crescita digitale 2014-2020*, Roma, 3 marzo 2015, p. 44 ss.. Consultabile in [http://www.agid.gov.it/sites/default/files/documentazione/strat\\_crescita\\_digit\\_3marzo\\_0.pdf](http://www.agid.gov.it/sites/default/files/documentazione/strat_crescita_digit_3marzo_0.pdf).

come sono stati delineati in questo lavoro, per poi utilizzarli per migliorare la partecipazione dei cittadini, l'efficienza dell'azione amministrativa e l'efficacia della decisione pubblica<sup>270</sup>.

Il legame tra valorizzazione del patrimonio informativo pubblico e *Big data* è un elemento che caratterizza anche analisi europee e internazionali, anch'esse finalizzate alla migliore decisione pubblica. Abbiamo già evidenziato come il concetto di data driven innovation sia stato introdotto proprio per esemplificare in poche parole come dalla raccolta e dall'organizzazione dei dati, dunque dalla valorizzazione del patrimonio informativo pubblico, si possano generare delle informazioni e da queste nuova, completa e originale conoscenza sulla realtà che l'amministrazione è chiamata a governare o sulla propria azione amministrativa, se siamo in presenza di prestazioni di servizi<sup>271</sup>. Una interessante conferma ci arriva anche da alcuni documenti e amministrazioni europee, come quella britannica, e da alcune amministrazioni extra-UE, come quella australiana e quella statunitense<sup>272</sup>. Tutti i documenti sottolineano e mettono in evidenza come solo con politiche di ammodernamento tecnologico e di ripensamento dell'organizzazione dei dati possono essere utilizzati i *Big data* per prendere delle decisioni amministrative o, come nel caso dell'amministrazione australiana, per migliorare l'efficienza dei servizi pubblici.

#### *4. Big data, amministrazioni pubbliche, cittadino: elementi per una nuova relazione cittadino-amministrazione?*

Dopo avere analizzato quali sono i rilievi giuridici che possono emergere dall'utilizzo dei *Big data* durante l'azione amministrativa, attraverso l'analisi delle dinamiche

---

270 Per le politiche sulle *smart cities* si v. Presidenza del Consiglio dei ministri, *Strategia per la crescita digitale 2014-2020*, Roma, 3 marzo 2015, p. 111 ss.. Consultabile in [http://www.agid.gov.it/sites/default/files/documentazione/strat\\_crescita\\_digit\\_3marzo\\_0.pdf](http://www.agid.gov.it/sites/default/files/documentazione/strat_crescita_digit_3marzo_0.pdf). La considerazione delle Smart cities come un incubatore di elaborazione delle strategie per valorizzare i *Big data* si v. anche OCSE, *Data-driven innovation. Big data for growth and well-being*, cit., p. 379 e ss.

271 Si v. diffusamente OCSE, *Data-driven innovation. Big data for growth and well-being*, cit..

272 Per approfondire la strategia australiana sui *Big data* si v. Dipartment of Finance, *The Australian Public Service Big Data Strategy: Improved Understanding through Enhanced Data-analytics Capability Strategy Report*, 2013; Per le considerazioni sul patrimonio informativo pubblico di quella statunitense si v. Executive Office of the President of the U.S.A., *Big data: seizing opportunities, preserving values*, Washington, Maggio 2014. Anche la Commissione europea sottolinea l'importanza e le opportunità legate alla valorizzazione del patrimonio informativo pubblico. Per approfondire COM (2014) 442, Comunicazione della Commissione europea al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *Verso una florida economia basata sui dati*; Anche il Parlamento britannico ha trattato il tema in House of Commons, Science and Tecnology Committee, *The Big data dilemma*, Fourth Report of session 2016-16.

procedimentali e delle tutele che si formano all'interno dei procedimenti tributari di accertamento standardizzato, e nell'organizzazione amministrativa, in particolare nell'organizzazione della gestione dei dati pubblici e privati in possesso delle pubbliche amministrazioni, è utile soffermarci su un altro profilo rilevante delle modifiche che i *Big data* possono apportare alle pubbliche amministrazioni, che assume un rilievo giuridico di non poco momento: il rapporto tra cittadino e amministrazione.

Il diluvio di dati, che investe e investirà sempre più le società contemporanee, tende ad accentrare il potere informativo in pochi soggetti, capaci di poter raccogliere, conservare ed elaborare i *Big data*. Questi soggetti sono innanzitutto le amministrazioni pubbliche, le quali sono in possesso di enormi quantità di dati pubblici e privati che utilizzano per le loro finalità istituzionali. Le amministrazioni pubbliche, come abbiamo potuto vedere in questo lavoro, se da un lato, non hanno ancora valorizzato a pieno il loro patrimonio conoscitivo, dall'altro, stanno cominciando ad adattare le loro capacità conoscitive alle caratteristiche del fenomeno dei *Big data* e a costruirsi un potere informativo fatto di concentrazione, elaborazione dei dati e produzione di nuova conoscenza.

Oltre alle amministrazioni pubbliche, è soprattutto nel settore privato che si sono formati soggetti che possiedono un elevato potere informativo. I provider del web (Google, Facebook, ect.) o le aziende che, nei più disparati settori, sono in possesso di grandi quantità di dati e sono capaci di elaborarli e utilizzarli (si pensi ad Amazon alla Microsoft o ad IBM), utilizzano i dati per migliorare la loro produttività, i loro profitti e la loro capacità di influenzare anche le decisioni pubbliche che le riguardano o che riguardano i propri settori.

La diffusione sempre maggiore delle tecnologie di elaborazione dei dati, le politiche di diffusione dei dati e la sua crescente centralità, però, come abbiamo già segnalato nel primo capitolo, potrebbe fare sorgere anche una terza tipologia di soggetti possessori e produttori di un potere informativo e conoscitivo: i cittadini. Alcuni studiosi hanno già sottolineato come i *Big data* possono mettere al centro la persona, in maniera del tutto inedita. Se, come abbiamo potuto vedere dagli esempi fatti nel corso di questo scritto, l'individuo è una delle maggiori fonti di produzione di *Big data*, in futuro potrebbe cambiare il rapporto tra la persona e i propri dati: l'individuo potrebbe avere un ruolo centrale nella gestione dei propri dati mai avuto precedentemente; potrebbe utilizzare i dati in modo tale da tutelare i propri interessi o pretendere il rispetto di diritti garantiti dalla legge<sup>273</sup>. Questa nuova prospettiva di maggiore consapevolezza del valore dei propri dati e delle possibilità di utilizzarli attivamente da parte dei cittadini potrebbe avere delle interessanti evoluzioni anche su un piano più generale: i cittadini potrebbero

---

273 V. MAYER SCHONBERGER – K. CUKIER, *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, cit.

utilizzare non solo i propri dati, ma anche i dati che è possibile reperire o che è possibile raccogliere nel web o tramite le regole di accessibilità pubbliche e private, e potrebbero utilizzarli non più soltanto per scopi personali, ma per scopi di interesse generali, per scopi collettivi.

La terza tipologia di soggetto - che detiene un più o meno intenso potere informativo e che potrebbe sorgere in maniera molto più graduale - è la cittadinanza attiva che utilizza i dati in proprio possesso, raccoglie i dati che le sono utili e gli elabora per tutelare interessi e diritti della collettività e, in determinati casi, ad apportare un contributo costruttivo alla decisione pubblica.

È utile allora riportare un esempio emblematico che va in questo senso, che non solo mette in rilievo quali sono le implicazioni giuridiche che emergono nella relazione cittadini-amministrazione nell'epoca dei *Big data*, ma delineano uno scenario esemplare di come si attegga il potere informativo in un momento in cui i dati e le informazioni che da essi si possono ricavare diventano centrali per prendere una decisione pubblica in maniera consapevole, informata e imparziale.

#### *4.1. Il caso dell'epatite C: come l'elaborazione e l'integrazione dei dati da parte dei cittadini influisce sulla decisione pubblica.*

Il caso che vale la pena di riportare è quello in merito alle politiche farmaceutiche del nostro Paese riguardanti i farmaci innovativi, in particolare riguardanti i nuovi farmaci che permetterebbero l'eradicazione dell'Epatite C. L'infezione da virus della epatite C (HCV) è la più comune causa di malattia cronica epatica diffusa nel mondo. Negli ultimi 3 anni si è assistito a un radicale cambiamento delle prospettive terapeutiche per i malati di epatite C grazie alla ricerca delle grandi case farmaceutiche e attraverso l'introduzione degli anti-virali diretti (direct-acting antiviral, DAA), i quali - seppure con percentuali diverse di successo in base al genotipo virale e all'entità del danno epatico - riuscirebbero a determinare in una elevata percentuale di casi l'eradicazione del virus dopo 3-6 mesi di trattamento. Dei farmaci fondamentali che, da un lato, si sono rivelati una risorsa fondamentale in una prospettiva di eradicazione totale del virus in Europa e conseguentemente in Italia e, dall'altro, hanno messo in crisi il sistema di *governance* della spesa farmaceutica di tutti gli Stati, dagli U.S.A. agli Stati europei.

In Italia il sistema di *governance* della spesa farmaceutica è caratterizzato da due elementi fondamentali - la definizione a priori di un tetto alla spesa farmaceutica, e il successivo meccanismo del *payback*. L'accesso all'innovazione farmaceutica avviene

permettendo all'azienda produttrice del farmaco riconosciuto come innovativo di non pagare la sua quota di *payback* in caso di sfioramento della spesa<sup>274</sup>.

In caso di innovazioni che richiedono una spesa rilevante, questo approccio, secondo alcuni studiosi, implica che il sistema possa dimostrarsi incapace di procedere con la spesa farmaceutica necessaria per accedere al nuovo farmaco in quanto il budget predeterminato di spesa farmaceutica risulta essere troppo rigido ed eccessivamente legato alla spesa sanitaria in generale<sup>275</sup>. Questa è stato proprio ciò che è accaduto con i nuovi farmaci per combattere il virus dell'epatite C, il cui utilizzo (50.000 euro di costo per terapia per milioni di persone) rischia di far fallire qualsiasi budget di bilancio dedicato alla spesa farmaceutica<sup>276</sup>.

Il governo in carica ha risposto con delle misure giudicate dagli studiosi molto importanti che, però, si sono rivelate ancora largamente inefficaci per garantire l'accesso ai nuovi farmaci.

La legge n. 190/2014 (Legge di stabilità 2015), infatti, nonostante abbia istituito nello stato di previsione del Ministero della salute, con l'art. 1, co. 593-598, un Fondo per il rimborso alle regioni per l'acquisto di medicinali innovativi, con lo scopo di riuscire ad acquistare i farmaci innovativi destinati alla cura dell'Epatite C, ha garantito risorse per il 2015 e per il 2016 che ammontano circa ad una cifra di un miliardo di euro, utile per garantire la cura a 50.000 persone su circa 2 milioni di malati di epatite C in Italia: una misura assolutamente insufficiente.

La vicenda, però, al di là della decisione del governo di istituire il Fondo per i farmaci innovativi, rileva per un punto fondamentale. Mancava un dato fondamentale per avviare una programmazione di eradicazione nazionale dell'epatite C complessiva e

---

274 Il meccanismo di ripiano del *pay back* consente, da un lato, l'erogazione di risorse economiche alle Regioni a sostegno della spesa farmaceutica di ciascuna, e dall'altro l'opportunità per le aziende farmaceutiche di effettuare le scelte sui prezzi dei loro farmaci, sulla base delle proprie strategie di intervento sul mercato. E' stato previsto nella legge finanziaria del 2007 e permette alle aziende farmaceutiche di chiedere all'Agenzia Italiana del Farmaco (AIFA) la sospensione della riduzione dei prezzi del 5%, a fronte del contestuale versamento in contanti (*pay back*) del relativo valore su appositi conti correnti individuati dalle Regioni.

275 Per approfondire il sistema di *governance* della spesa farmaceutica e la distribuzione della competenza tra Stato e Regioni si v. G. MASSARI, *La sicurezza dei farmaci in Italia, tra vincoli sopranazionali e riparto interno di competenze*, in *Ist. Fed.*, 2015, fasc. 1, p. 141 ss.; TAR Lazio - Roma sez. I 2 dicembre 2014, n. 1216, con nota di R. PARDOLESI, *Concorrenza sul mercato: produzione di farmaci, collocazione a carico dello stato ed esclusione di intese restrittive volte a promuovere l'inserimento nel mercato del più costoso*, in *Foro it.*, 2015, fasc. 1, p. 38 ss.; G. DEMURO, *La leale collaborazione nella differenziazione della spesa farmaceutica*, in *Le Regioni*, 2012, fasc. 1-2, p. 432 ss.; L. CUOCOLO, *I rapporti Stato-Regioni in materia farmaceutica, tra diritti costituzionali e contenimento della spesa*, in *Quad. reg.*, 2008, fasc. 1, p. 33 ss.;

276 I numeri sulle coperture previste dal governo si possono reperire in F. SPANDONARO, *Le politiche farmaceutiche e i fondi per l'innovazione: come garantire la sostenibilità*, in [http://www.quotidianosanita.it/studi-e-analisi/articolo.php?articolo\\_id=20939](http://www.quotidianosanita.it/studi-e-analisi/articolo.php?articolo_id=20939).

organizzata: una stima realistica del numero dei pazienti eleggibili ad un trattamento antivirale. Un dato che i cittadini, elaborando una grande quantità di dati, sono riusciti a ricostruire e a fornire all'amministrazione.

L'Associazione EpaC Onlus, un'associazione che si occupa da tempo di questi temi, ha avviato una ricerca lunga e complessa, durata un anno, che l'ha portata a raccogliere i dati sulle esenzioni per patologia da epatite C direttamente dalle singole Regioni e/o ASL, e ad integrarli con i dati su altri aspetti come il numero dei pazienti malati con esenzioni diverse, il numero dei pazienti guariti ma con esenzione, ed dati su altri elementi che concorrono a formare una stima complessiva e realistica dei pazienti eleggibili ad un trattamento antivirale. Lo studio ha portato a rilevare come sono circa 160/180mila i pazienti con HCV eleggibili ad un trattamento antivirale con i farmaci innovativi, cifra di gran lunga inferiore rispetto a quella sinora stimata in Italia dal governo. Contestualmente, per capire l'impatto economico sul Servizio Sanitario Nazionale, l'associazione ha chiesto all'Università di Roma 'Tor Vergata' (CEIS-EEHTA) di stimare l'impatto economico usando questa popolazione e considerando i livelli di prezzo decrescente negli anni dei nuovi farmaci. Nell'ipotesi in cui si decidesse di trattare 30.000 pazienti ogni anno, al 2022 avremmo una spesa complessiva di 2.5/2.8 miliardi di euro (con una spesa media per paziente trattato di circa € 15.000). Considerando un recente studio che ha stimato una riduzione di costo attribuibile ai livelli di efficacia dei nuovi farmaci anti-HCV pari a € 12.500 per paziente trattato già a partire dal 2018, l'investimento proposto sembra avere tutti i criteri di una sostenibilità economico-finanziaria<sup>277</sup>. Una proposta alternativa, complessiva e ragionata che è stata presa in considerazione anche dal Ministero della Salute stesso.

Una vicenda che aveva fatto emergere anche come la mancanza di una reale conoscenza da parte del governo della situazione complessiva del numero dei malati che avrebbero dovuto usufruire dei nuovi farmaci abbia portato ad una "cattura" dell'Agenzia Italiana del Farmaco (AIFA) da parte dell'azienda farmaceutica che stava trattando per la vendita del farmaco in Italia, tanto da portare a ripetute interrogazioni parlamentari sul punto<sup>278</sup>.

#### *4.2. L'elaborazione dei Big data da parte dei cittadini come nuovo strumento di equilibrio del potere informativo.*

---

277 MENNINI et al., *Disponibilità a pagare e innovazione: il caso dei farmaci anti-HCV nel Sistema Sanitario Italiano*, GRHTA, 2015, 2(2), p. 69.

278 Per approfondire le relazioni del Forum parlamentare si consulti [http://www.sosfegato.it/camo/onlus/all/I\\_Forum\\_parlamentare\\_hcv.pdf](http://www.sosfegato.it/camo/onlus/all/I_Forum_parlamentare_hcv.pdf)

La vicenda che abbiamo brevemente illustrato fa emergere due aspetti importati, utili per il nostro discorso. Uno essenzialmente generale che, come dicevamo in precedenza, mette in evidenza le dinamiche che vengono a formarsi tra i soggetti che detengono il potere informativo e l'altro molto più utile in quanto stimolano interessanti riflessioni giuridiche in merito al rapporto tra cittadino e amministrazione.

In merito al primo aspetto la vicenda delle politiche per la diffusione di farmaci innovativi utili per curare l'epatite C, fa emergere come l'incapacità o, in questo caso, la mancanza di volontà politica e amministrativa di produrre informazioni dai dati che si posseggono, ha portato l'amministrazione sanitaria italiana a essere "catturata" da un soggetto, la casa farmaceutica, in quanto non sembrava possedere tutte le informazioni per comprendere pienamente la realtà su cui doveva intervenire. Ci troviamo di fronte ad un classico esempio di come l'informazione e la conoscenza che essa genera possa portare vantaggi o svantaggi all'uno o all'altro soggetto che detiene il potere informativo. Una dinamica a cui si è aggiunta la capacità dei cittadini organizzati di incrinare l'equilibrio (o il disequilibrio) che si è venuto a creare, utilizzando allo stesso modo l'elaborazione dei dati e delle informazioni.

La raccolta dei dati, la capacità di integrarli e fonderli e la capacità di produrre delle informazioni da essi, si rivela come lo strumento fondamentale per stabilire equilibri o disequilibri informativi tra i soggetti pubblici e privati durante la definizione di qualsiasi politica pubblica.

In merito al secondo aspetto, vale la pena sottolineare come l'iniziativa di cittadini organizzati di raccogliere dei dati in possesso delle Regioni e che l'amministrazione pubblica a tutti i livelli non utilizza e non organizza e di integrarli con altri dati rilevanti per le finalità che si sono dati (tentare di comprendere il numero di pazienti eleggibili per la cura) sta cambiando una politica pubblica (quella del ministero della sanità in merito alla spesa farmaceutica innovativa) fra le più rilevanti del secolo. Una vicenda che fa sorgere interrogativi interessanti sul ruolo informativo che i cittadini possono avere, non solo in quanto portatori di interessi personali o collettivi da fare valere attraverso l'elaborazione dei dati, ma anche in quanto portatori di nuove informazioni e nuova conoscenza dei fenomeni che l'amministrazione pubblica non conosce o non produce da sola.

La vicenda che abbiamo trattato non è l'unica, in quanto numerosi sono gli esempi in cui i cittadini, singoli o riuniti in associazioni o comitati, cominciano a rapportarsi con l'amministrazione in maniera attiva e consapevole utilizzando gli strumenti conoscitivi a loro disposizione, sia grazie alla possibilità, oramai tendenzialmente consolidata nel nostro ordinamento, di riutilizzare i dati pubblici, sia grazie alle possibilità di reperire dati che l'amministrazione non possiede. Questa prospettiva non può che essere

potenziata dalla diffusione dei *Big data*, considerando che il settore privato ha saputo sfruttare prima e meglio del settore pubblico i suoi strumenti e le sue potenzialità.

La cittadinanza amministrativa nell'era dei *Big data*, potrebbe, quindi, contribuire a realizzare un differente rapporto tra cittadini e amministrazione, vicino a quello prospettato poc'anzi: un'amministrazione che garantisce diritti informativi con un approccio più ampio possibile, certa di avere un importante ritorno informativo per potere decidere in maniera più consapevole e una cittadinanza che utilizza i propri diritti informativi, non solo per tutelare le proprie situazioni giuridiche, ma anche per arricchire la conoscenza pubblica.

## Capitolo 4

### Una nuova funzione conoscitiva per le amministrazioni pubbliche

Sommario: 1. I *Big data* nelle amministrazioni pubbliche; 2. Le motivazioni di una rinnovata disciplina dell'azione e dell'organizzazione amministrativa.; 3. Principio di legalità e tipizzazione del potere conoscitivo delle amministrazioni.; 4. Una nuova funzione conoscitiva delle pubbliche amministrazioni.; 5. *Big data* e *Open data*: differenze concettuali e nuovi equilibri tra trasparenza e *privacy*.; 6. Una nuova cittadinanza amministrativa con i *Big data*

#### *1. I Big data nelle amministrazioni pubbliche.*

L'esplosione dei dati, le loro nuove caratteristiche, l'eterogeneità delle fonti di produzione e, in particolar modo, i nuovi fenomeni e i nuovi strumenti conoscitivi che si sono strutturati per la loro elaborazione, in sintesi, il fenomeno dei *Big data* e la sua diffusione nella società, ha fatto emergere dei rilievi giuridici molto rilevanti. Abbiamo visto nel secondo capitolo come le analisi principali si siano soffermate sui rischi e su come difendersi dal fenomeno, sottolineando come la concentrazione di potere informativo nelle mani di pochi soggetti e la loro capacità di elaborare i dati con tecnologie che sfruttano in maniera nuova e più incisiva algoritmi matematici e inferenze statistiche abbiano messo in crisi molti aspetti del diritto antitrust, abbiano sostanzialmente paralizzato la disciplina sulla protezione dei dati personali e abbiano fatto emergere anche dei profili collettivi rilevanti del rapporto tra i *Big data*, da un lato, e la *privacy* e le libertà fondamentali dei cittadini, dall'altro: si pensi al nuovo potere di controllo sociale che l'elaborazione dei *Big data* sta costruendo in maniera sempre più penetrante a vantaggio delle grandi aziende che utilizzano i dati e che alcuni studiosi hanno paragonato a quello dello Stato; si pensi al potenziamento del controllo del potere pubblico sui cittadini attraverso l'utilizzo dei *Big data* per tutelare la sicurezza pubblica e per combattere la criminalità che rischia di ledere pesantemente le libertà fondamentali dei cittadini.

Alle analisi che si sono occupate di come difendersi dai *Big data*, si affiancano alcune analisi che, invece, cominciano a mettere in rilievo anche come utilizzare i *Big data*, in particolare nei processi decisionali pubblici. Anche se queste analisi non oscurano gli aspetti negativi che un disequilibrio di potere informativo tra i soggetti pubblici e privati che si confrontano durante i processi decisionali pubblici, possono generare (si pensi agli esempi che abbiamo riportato nel corso del lavoro sui casi di “cattura” del decisore pubblico o di tentativi di influenza da parte di soggetti privati con un maggiore potere

informativo sul decisore pubblico), queste analisi risultano essere rilevanti in quanto mettono in evidenza come l'elaborazione dei *Big data*, le correlazioni e le analisi predittive, la capacità di questi nuovi strumenti conoscitivi di creare nuove informazioni e nuova conoscenza dai dati, la possibilità di avere una migliore e più completa comprensione della realtà, danno agli organi politici tutti gli elementi per modificare le opzioni regolatorie (*hard law*, *soft law* o non regolare affatto) o le scelte regolative su una determinata materia.

L'utilizzo dei *Big data* per la decisione pubblica, però, è da intendersi in senso ampio. È vero che le analisi che abbiamo considerato in questo lavoro si fermano alla possibile utilità dell'elaborazione dei *Big data* per prendere delle decisioni essenzialmente politiche, ma – nonostante non ci siano ancora studi rilevanti nella dottrina amministrativistica – abbiamo potuto verificare nel capitolo precedente come l'elaborazione di dati con funzioni matematiche e inferenze statistiche (analisi dei dati molto vicine a quelle che si fanno con i *Big data*) vengano già utilizzate dalle amministrazioni pubbliche per prendere delle decisioni amministrative e per svolgere le funzioni che ad esse vengono affidate dalla legge. L'amministrazione finanziaria, infatti, durante le sue funzioni di controllo e in particolare durante i procedimenti di accertamento standardizzato, elabora i dati che sono in suo possesso attraverso degli strumenti matematici e statistici (gli studi di settore e il redditometro) per accertare il corretto adempimento da parte dei contribuenti degli obblighi fiscali.

Nella dottrina tributaria l'utilizzo istruttorio di questi strumenti ha portato non solo ad una riflessione annosa sui punti problematici più importanti, ma anche a dei punti fermi su quegli stessi problemi, confermati anche dalla giurisprudenza tributaria: si pensi alla diffusa considerazione dei risultati di questi strumenti come presunzioni semplici, con le relative ricadute in termini processuali; si pensi all'obbligo del contraddittorio procedimentale con il contribuente preventivo rispetto all'avviso di accertamento; si pensi alla necessità, più volte sottolineata dalla Cassazione, di utilizzare questi strumenti nel rispetto dei principi di ragionevolezza e di proporzionalità.

Le amministrazioni pubbliche non stanno utilizzando, di fatto, i *Big data* solo in determinati procedimenti, se vogliamo anche marginali e settoriali, ma sta cominciando anche a organizzare la gestione dei propri dati in maniera tale da potere raccogliere conservare ed elaborare al meglio i dati con le caratteristiche dei *Big data*.

## *2. Le motivazioni di una rinnovata disciplina dell'azione e dell'organizzazione amministrativa.*

Le tipologie di utilizzo dei Big data che fino ad ora sono state portate avanti evidenziano anche aspetti e prospettive che non sono state ancora indagate a pieno e che qui tenteremo di mettere in fila in sommi capi per poi ritornarci nei capitoli successivi. Se le implicazioni che abbiamo evidenziato prima sono implicazioni molto evidenti e tra le più indagate dagli studiosi, esistono delle implicazioni giuridiche che ricadono sull'azione e sull'organizzazione delle amministrazioni pubbliche che meritano di essere elencate ed analizzate. Molti degli esempi che sono stati riportati dimostrano come i Big data stanno modificando lentamente, ma sostanzialmente il modo di prendere le decisioni delle pubbliche amministrazioni.

Le correlazioni e i rinnovati metodi statistici e matematici di elaborazione dei Big data si stanno rivelando come nuovi strumenti conoscitivi nelle mani delle pubbliche amministrazioni attraverso i quali queste riescono ad acquisire conoscenze e punti di vista differenti ed originali. Abbiamo già detto, inoltre, come le correlazioni, appunto perché scevre da qualsiasi impostazione preliminare, da qualsiasi ipotesi iniziale, siano capaci di evidenziare relazioni tra variabili, originali, poco intuibili, che mettono i fenomeni che vengono analizzati sotto una luce differente. Abbiamo già spiegato, inoltre, come le correlazioni non vogliono legare quelle variabili con un nesso di causalità, ma sono utili per mettere in evidenza una semplice relazione statisticamente rilevante.

L'elaborazione dei Big data, poi, ha una forte componente predittiva: la capacità di poter analizzare una notevole quantità di dati, talmente tanto notevole da abbassare sostanzialmente il rischio statistico - almeno a breve-medio termine - e prevedere probabilisticamente le dinamiche che riguardano un determinato fenomeno. Questi elementi conoscitivi, considerando i numerosi esempi riportati nei paragrafi precedenti, hanno migliorato la capacità conoscitiva e conseguentemente quella decisionale delle pubbliche amministrazioni, hanno permesso a queste ultime di governare con i Big data.

Governare con i “*Big data*” significa sapere elaborare i dati di fonte pubblica e privata posseduti dalle pubbliche amministrazioni - producendo ulteriori informazioni utili allo svolgimento delle proprie funzioni - ed utilizzare i dati in maniera tale da migliorare la conoscenza delle situazioni di fatto su cui l'amministrazione deve intervenire, la rapidità e l'efficacia dell'intervento pubblico - sfruttando le capacità predittive di queste tipologie di analisi - e l'efficienza e la buona gestione dei servizi pubblici.

Utile è allora comprendere in questo paragrafo, innanzitutto, quali sono le implicazioni giuridiche che un fenomeno di questo genere apre sulla disciplina dell'azione amministrativa, metterle in fila, per poi cercare di capire, nel prossimo capitolo, quanto la sua disciplina positiva sia in grado di rapportarsi al fenomeno dei Big data, riuscendo a interpretarne le sfide. Le implicazioni giuridiche sulla decisione amministrativa e sulla sua formazione costituiscono l'aspetto più interessante della questione, in quanto incidono profondamente non solo sulla disciplina positiva del procedimento amministrativo, in particolare quella sull'istruttoria, ma anche sulla natura dell'attività conoscitiva delle pubbliche amministrazioni e sul concreto esercizio della discrezionalità amministrativa.

L'elaborazione dei Big data, come dimostrano gli esempi che abbiamo riportato nei paragrafi precedenti, apre alle amministrazioni pubbliche prospettive conoscitive nuove sulla realtà che le circonda e che sono chiamate a governare: la possibilità di raccogliere tutte le informazioni possibili sui fatti e sui fenomeni che si vogliono indagare, mette l'amministrazione nelle condizioni di conoscere davvero i fatti, gli interessi, i bisogni della comunità che governa, in maniera sempre più autonoma e sempre più imparziale; le analisi probabilistiche e tendenzialmente predittive sui fenomeni e la capacità, attraverso le correlazioni, di legare aspetti in via del tutto originale e molte volte impensabile, dà alle amministrazioni la possibilità di migliorare ulteriormente la conoscenza della realtà, ma anche di velocizzare il processo decisionale e migliorare l'efficacia della sua azione amministrativa, permettendole di capire molto prima le tendenze dei fenomeni che vuole governare. In termini più strettamente giuridici le possibilità conoscitive che la raccolta e l'elaborazione dei Big data mettono a disposizione dell'amministrazione rendono ancora più evidente come sia necessario ritornare a riflettere sull'attività conoscitiva delle amministrazioni e sulle sue prospettive evolutive in un contesto di questo genere. Le potenzialità conoscitive dei Big data non possono essere colte a pieno se si considera l'attività conoscitiva delle amministrazioni come un'attività meramente strumentale ad un determinato procedimento amministrativo piuttosto che come una funzione autonoma, con le sue regole procedurali e organizzative. In realtà non si potrebbero cogliere queste potenzialità neanche se si slegasse la dimensione legata all'azione amministrativa della funzione conoscitiva dalla dimensione più organizzativa, legata alla gestione dei dati pubblici. Le due dimensioni sono strettamente connesse e, cercheremo di sottolinearlo nel corso di questo lavoro.

Le potenzialità conoscitive dei Big data, inoltre - grazie alle sue caratteristiche, ai suoi strumenti e alla capacità di sapere cogliere, almeno potenzialmente, la realtà nella sua totalità e complessità - ci portano a riflettere anche su come questa nuova conoscenza

influisce sulla disciplina del procedimento, in particolare quella inerente l'istruttoria, e della decisione amministrativa.

I principi che governano l'istruttoria procedimentale, il principio di informalità e il principio (o metodo) inquisitorio, si confrontano con degli strumenti di reperimento degli elementi conoscitivi del tutto atipici, i quali, da un lato, rischiano di modificare profondamente alcuni aspetti (nel caso del principio di informalità) dell'istruttoria procedimentale, dall'altro, fanno esplodere le potenzialità di altri aspetti (nel caso del principio inquisitorio), modificando la declinazione e l'intensità dei principi che la guidano. La mancanza di qualsiasi tipizzazione degli strumenti per acquisire gli elementi conoscitivi nella legge sul procedimento, certamente non ostacola l'utilizzo di strumenti atipici, come quelli in esame, ma allo stesso tempo questi non corrispondono a quelli che sovente l'amministrazione utilizza, siano questi atti di autorità (acquisizioni di scienza, acclaramenti, richieste documentali, accessi interamministrativi, richieste di pareri o valutazioni tecniche, studi e rapporti) o classici moduli conoscitivi amministrativi (ispezioni, inchieste, perizie).

### *3. Principio di legalità e tipizzazione del potere conoscitivo delle amministrazioni.*

I dati e le informazioni prodotte dai Big data possono essere, infatti, il presupposto di fatto che legittima l'avvio di un procedimento amministrativo tradizionale. In questo caso il problema giuridico più rilevante è comprendere come riversare le informazioni e i dati prodotti con le nuove tecnologie e i nuovi metodi di elaborazione dei Big data all'interno delle fasi procedurali e della successiva motivazione, come rendere questi dati e queste informazioni giuridicamente rilevanti all'interno del procedimento e come fare rispettare i principio di affidabilità e di legalità che reggono tutto il procedimento amministrativo.

I dati e le informazioni frutto dell'elaborazione e dell'analisi dei Big data, però, possono diventare anche il nucleo centrale dell'istruttoria procedimentale, possono, non solo legittimare il potere discrezionale che agisce secondo le tradizionali regole procedurali, ma fondare il potere stesso e la decisione amministrativa.

Si pensi agli effetti che questa tipologia di utilizzo dei Big data può avere nei procedimenti di accertamento e di controllo: l'amministrazione fiscale commina una sanzione a Tizio sulla base di un'analisi dei Big data che rileva uno scostamento medio negativo (ricordiamo, frutto di una correlazione, non di un'analisi causale) dei suoi pagamenti. Oppure si pensi al caso in cui un'amministrazione comunale decidesse di

chiudere il traffico o diminuirne l'intensità sulla base di un'analisi dei Big data che prevedesse, con una probabilità alta, del maltempo sulla zona, per poi dovere ritornare sulla sua decisione, provocando dei disagi ai cittadini. In questo caso il problema giuridico diventa maggiormente rilevante: può il potere amministrativo avere una base di legittimazione di questo tipo? Fino a che punto possono essere utilizzati i Big data all'interno del procedimento amministrativo?

Quest'ultimo esempio dimostra come i Big data non solo rendono l'attività conoscitiva delle amministrazioni pubbliche ancora più autonoma, rendendola una funzione a se stante, ma trasforma la dimensione dinamica legata all'azione amministrativa della funzione conoscitiva in un vero e proprio potere conoscitivo perché ha un nucleo di potenzialità di compressione dei diritti molto pronunciato. Un potere conoscitivo che modifica anche la discrezionalità amministrativa appiattendola totalmente al dato tecnico delle elaborazioni sui Big data, con il rischio di legittimare una discrezionalità senza nessuna garanzia per i cittadini.

L'utilizzo dei Big data per la decisione amministrativa, infatti, potrebbe minare, in determinate occasioni, alcune garanzie procedurali (si pensi ai diritti di partecipazione) e la ratio profonda, non solo della fase istruttoria, ma di tutto il procedimento amministrativo, di garanzia e affidamento per i cittadini.

I Big data, però, se da un lato ci costringono a riflettere sulla bontà delle sue implicazioni sulla disciplina procedimentale, sull'attività/potere conoscitivo delle pubbliche amministrazioni e sulle trasformazioni della discrezionalità amministrativa, dall'altro, potenzia qui in maniera del tutto positiva la possibilità, per le amministrazioni di avvicinarsi alla verità materiale dei fatti e delle situazioni che le amministrazioni analizzano. I Big data fanno esplodere le potenzialità del principio inquisitorio che regge l'istruttoria procedimentale e danno all'amministrazione degli strumenti che sono in grado davvero di permetterle di conoscere la realtà in tutta la sua complessità e anche nella sua interezza. Con i Big data la certezza pubblica non è solo tale perché fondata sull'autorità del soggetto che la afferma (l'amministrazione pubblica) e la decisione amministrativa non è più una scelta adeguatamente giustificata o la scelta sufficiente, ma la certezza pubblica è tale perché fondata sull'essenza oggettiva dei fenomeni (pur negli inevitabili margini di errore dello strumento e dell'interpretazione dell'uomo), sulla realtà così com'è e la decisione amministrativa si avvicina moltissimo alla decisione ottimale, alla migliore delle decisioni possibili. Il principio di verità materiale, tendenzialmente recessivo negli ultimi anni e riportato in auge in dottrina in pochi casi, può tornare ad essere una prospettiva anche nel senso di permettere all'amministrazione di essere fino in fondo un'autorità razionale.

Una prospettiva che non può essere ignorata quando si parla di Big data in relazione all'azione amministrativa - dopo avere posto in maniera molto rapida le implicazioni giuridiche sull'amministrazione - è quella degli effetti esterni all'amministrazione: gli effetti sui cittadini e il rapporto tra amministrazione e giudice.

Una prima problematica inerentemente gli effetti sui cittadini è legata al rispetto della disciplina sulla tutela dei dati personali dei cittadini. Le tipologie di informazioni che è possibile archiviare con i Big data, che possono riguardare aspetti fortemente personali (come i dati biometrici o i dati sulle abitudini e sul comportamento), la loro quantità, i metadati - che permettono sia alle amministrazioni sia ai privati, quando siamo in presenza di politiche di open data, di elaborarli e ricavarne nuove e ulteriori informazioni personali, anche molto dettagliate - ci costringono a ripensare a quali dati o banche dati aprire ai cittadini e ai privati - e quindi a quale rapporto bisogna costruire tra Big data, open data e privacy - o a quali dati o informazioni pubblicare nei siti istituzionali delle pubbliche amministrazioni - e quindi a quale rapporto si deve instaurare tra Big data, trasparenza amministrativa e privacy. Questo approccio è l'unico che ci permette di recuperare la tutela dei dati personali nel rapporto tra cittadini e pubbliche amministrazioni, in quanto l'utilizzo dei Big data sta rendendo recessivo il principio-cardine della tutela della riservatezza dei dati personali: il principio del consenso informato preventivo. Il continuo riutilizzo dei Big data, che contengono anche dati personali, sta paralizzando la disciplina della privacy nel nostro ordinamento, in Europa e negli Stati Uniti in quanto gli strumenti solitamente utilizzati per garantirla possono essere sistematicamente superati con le elaborazioni dei Big data. Il divieto di indicizzazione e l'anonimizzazione, i due maggiori strumenti di garanzia della privacy, sono del tutto inefficaci a garantire la protezione dei dati personali dei cittadini.

La tutela della privacy quando i Big data vengono utilizzate dalle amministrazioni pubbliche, come già abbiamo rilevato nel paragrafo precedente, non è più soltanto un problema individuale, ma è sempre più un problema collettivo. L'elaborazione dei Big data può portare alla classificazione e alla profilazione dei cittadini, dividendoli in riferimento alle più svariate caratteristiche sociali o personali, in maniera più accurata rispetto al passato. Questa impostazione prescinde da qualsiasi azione individuale e qualsiasi forma di responsabilità personale: si potrebbe essere sanzionati non per qualcosa verificata concretamente, ma per scostamenti dal comportamento medio del gruppo sociale in cui l'individuo è stato inserito. Questa situazione - oltre a portare qualche studioso ad affermare che sia necessario discutere di privacy e Big data anche tenendo presente l'impatto sociale del fenomeno, quindi slegando la tutela della privacy dalla tutela del singolo dato personale - rischia di vanificare, almeno nel nostro ordinamento, alcuni aspetti della cittadinanza amministrativa. Se la cittadinanza amministrativa è un insieme di diritti e di doveri di solidarietà dell'individuo, all'interno

di un contesto in cui si sta perdendo progressivamente il paradigma dell'amministrazione come soggetto autoritario in favore di uno in cui è l'individuo ad avere una maggiore centralità e l'amministrazione diventa veicolo per la sua realizzazione all'interno della comunità, queste derive rischiano di annichilire queste tendenze a favore di un ritorno di un'amministrazione-autorità più forte, più moderna e più invasiva della precedente.

#### *4. Una nuova funzione conoscitiva delle pubbliche amministrazioni.*

Oltre ai rischi per il rapporto cittadini-amministrazione che provengono dall'utilizzo dei Big data da parte dell'amministrazione, i Big data possono rappresentare, dall'altro lato, un'opportunità per migliorare il controllo e la partecipazione democratica dei cittadini. La repentina diffusione delle tecnologie dei Big data e i diritti di riutilizzo dei dati e delle informazioni pubbliche diffuse dalle amministrazioni e contenute nelle banche dati possono costituire gli strumenti per la costruzione di un contropotere conoscitivo nelle mani dei cittadini. Numerosi sono gli esempi in cui i cittadini, singoli o riuniti in associazioni o comitati, cominciano a rapportarsi con l'amministrazione in maniera attiva e consapevole utilizzando gli strumenti conoscitivi a loro disposizione, sia grazie alla possibilità, oramai tendenzialmente consolidata nel nostro ordinamento, di riutilizzare i dati pubblici, sia grazie alle possibilità di reperire dati che l'amministrazione non possiede. Questa prospettiva non può che essere potenziata dalla diffusione dei Big data, considerando che il settore privato ha saputo sfruttare prima e meglio del settore pubblico i suoi strumenti e le sue potenzialità. L'obiezione che può essere fatta subito è legata alla considerazione che oggi le tecnologie che elaborano i Big data sono a disposizione di pochi soggetti privati, ma, in futuro, anche grazie ad una maggiore diffusione delle tecnologie che elaborano i Big data dovuta ad una loro più ampia accessibilità economica, la situazione prospettata potrebbe cambiare.

La cittadinanza amministrativa nell'era dei Big data, potrebbe, quindi, contribuire a realizzare un differente rapporto tra cittadini e amministrazione, vicino a quello prospettato poc'anzi: un'amministrazione che garantisce diritti informativi con un approccio più ampio possibile, certa di avere un importante ritorno informativo per potere decidere in maniera più consapevole e una cittadinanza che utilizza i propri diritti informativi, non solo per tutelare le proprie situazioni giuridiche, ma anche per arricchire la conoscenza pubblica.

Gli effetti che l'utilizzo dei Big data nell'azione amministrativa producono fuori dall'amministrazione non si limitano a modificare il rapporto tra amministrazione e

cittadini, in positivo e in negativo che sia, ma anche il rapporto tra giudice e amministrazione, in particolare tra decisione amministrativa e giudice.

La questione rilevante nel rapporto tra giudice e decisione amministrativa è legata alla possibilità da parte del primo di raggiungere un grado cognitivo adeguato sui fatti che gli si presentano in giudizio, soprattutto quando essi sono fatti complessi, questioni tecniche e legati al tema della discrezionalità tecnica. Di fronte ad un'amministrazione che governa con i Big data, il giudice dovrebbe poter essere in grado di conoscere, rispettando il limite del merito, i criteri, la logicità delle valutazioni tecniche, deve essere in grado di poter contrastare gli abusi che le amministrazioni possono mettere in atto utilizzando i Big data. I poteri cognitivi del giudice amministrativo devono andare di pari passo con quelli dell'amministrazione, tanto più se ci troviamo di fronte a casi in cui si è in presenza di un vero potere conoscitivo, il quale può potenzialmente comprimere i diritti dei cittadini. Per contrastare il potere conoscitivo delle pubbliche amministrazioni che utilizzano i Big data, non ci si può porre il tema del potere cognitivo del giudice, anche se qui lo tratteremo superficialmente poiché meriterebbe altra attenzione e altro spazio.

Le implicazioni sull'azione amministrativa e sulle sue numerose conseguenze non sono le uniche ad avere rilevanza nell'analisi dei Big data e delle sue implicazioni giuridiche sull'amministrazione pubblica. Governare con i Big data presuppone, infatti, sapere governare i Big data. Governare i Big data significa sapere gestire i dati pubblici e privati in proprio possesso e organizzare la propria amministrazione (dal punto di vista tecnologico, normativo e tecnico) in maniera tale da poter utilizzare effettivamente i Big data durante l'azione amministrativa.

Da questa breve definizione si evince che in questa parte del lavoro ci interrogheremo su come le amministrazioni pubbliche possono governare i Big data, non rimanendo confinati nella gestione dei dati pubblici e privati che possiedono quelle determinate caratteristiche particolari che li rendono Big data, ma adottando un'impostazione più larga. I dati pubblici e privati in possesso delle amministrazioni pubbliche a cui noi faremo riferimento non saranno soltanto quelli che hanno le caratteristiche dei Big data, ma, in generale, a tutti i dati che le amministrazioni hanno nella loro disponibilità: faremo riferimento a tutto il patrimonio informativo pubblico, composto dai dati tradizionali e dai “nuovi” dati.

Governare i Big data significa, innanzitutto, dotarsi di regole e di strumenti tecnici per riuscire a produrli, conservarli ed elaborarli. Nel corso del capitolo abbiamo sottolineato come i Big data siano dei dati con delle particolari caratteristiche (volume, velocità e varietà) che, in parte, li differenziano dai dati che tradizionalmente le pubbliche amministrazioni sono abituate a raccogliere, a conservare e a utilizzare. La possibilità

che questi possano essere prodotti, archiviati ed elaborati insieme ai dati tradizionali, modifica in parte le caratteristiche del patrimonio informativo delle amministrazioni, rendendolo non più soltanto un insieme di informazioni o di dati, ma anche di metadati, di dati grezzi, utili per produrre nuove informazioni attraverso la loro elaborazione e la loro analisi, ma che non sono immediatamente fruibili. Questo aspetto fa sorgere una prima problematica giuridica rilevante, che attiene alla formazione e alla gestione del patrimonio informativo pubblico, in particolare alla disciplina del dato e alla formazione delle banche dati.

Per quanto riguarda la disciplina del dato, si pensi all'impatto che i Big data possono avere sulla disciplina della qualità dei dati pubblici. Nei paragrafi precedenti abbiamo visto come la tendenza all'estrema valorizzazione dei dati, anche se di cattiva qualità, costituisca uno dei cambiamenti culturali maggiormente rilevanti dei Big data. Un cambiamento che porterebbe le amministrazioni, così come è avvenuto in parte negli esempi riportati e come avviene nel settore privato, a raccogliere tutti i dati possibili su un determinato fenomeno, anche dati di cattiva qualità, per sfruttare a pieno le metodologie di elaborazione dei Big data e per prendere la migliore decisione possibile. In quest'ottica l'amministrazione si troverebbe a conservare dei dati di cattiva qualità che la disciplina attuale obbliga a mettere a disposizione di tutte le amministrazioni, in funzione di cooperazione interamministrativa, e a disposizione dei cittadini, in funzione di trasparenza e di riutilizzo. Il primo problema a cui ci troviamo di fronte è, dunque, il problema della qualità dei dati pubblici e della sua funzione di garanzia sia per le altre amministrazioni, sia per i cittadini. Le domande che ci si possono porre sono molteplici: quanta cattiva qualità dei dati può essere accettabile all'interno del patrimonio informativo pubblico? Come cambia la disciplina della qualità dei dati pubblici? I principi di qualità dei dati individuati nel Codice dell'amministrazione digitale (CAD) sono ancora capaci di interpretare le nuove necessità di produzione e conservazione del patrimonio informativo pubblico?

Oltre alla disciplina della qualità dei dati, il fenomeno di Big data ha delle ricadute importanti sulla banche dati.

I Big data, come abbiamo descritto in precedenza, hanno reso inadeguate le tecnologie che supportavano le banche dati tradizionali, cioè quelle relazionali, le quali non sono più in grado di archiviare i dati che hanno le caratteristiche dei Big data. L'eterogeneità delle fonti e delle forme, la loro enorme quantità, la crescita esponenziale della loro produzione costringono le amministrazioni a ricorrere a nuovi strumenti tecnologici che, come abbiamo visto in precedenza, siano in grado di archiviare e produrre i Big data senza eccessiva difficoltà. La possibilità di utilizzare i Big data, quindi, passa dalla capacità delle amministrazioni pubbliche di sapere rinnovare i loro strumenti tecnologici

di conservazione adeguandoli alle tecnologie dei Big data. Le amministrazioni possono farlo, rinnovando i propri software, adeguandoli al livello tecnologico necessario oppure affidarsi a nuove metodologie di archiviazione svincolate dal possesso materiale del software, come i servizi di cloud computing. Nel primo caso le implicazioni sono inerenti le politiche di promozione per il rinnovamento della tecnologie all'interno delle amministrazioni. Le domande che ci si deve porre riguardano la valutazione sull'adeguatezza delle politiche sulla digitalizzazione e sullo sviluppo delle ICT. Nel secondo caso - agli aspetti inerenti le politiche pubbliche che abbiamo richiamato prima, poiché esse ingloberebbero anche eventuali politiche di promozione del cloud, o alle prassi che sono state attuate nei vari contesti – bisognerebbe analizzare anche alcune implicazioni giuridiche di non poco momento. I servizi di cloud computing sono offerti, infatti, da compagnie private, che mettono a disposizione di coloro che lo richiedono un servizio on-line (a pagamento, di solito) di archiviazione dei dati. Le aziende che forniscono questo servizio sul mercato sono, ovviamente, proprietarie delle infrastrutture all'interno delle quali confluiscono i dati e hanno la possibilità di fare accedere al servizio o di interromperlo. Queste dinamiche meritano di essere analizzate quando si tratta di dati pubblici ed è utile porsi delle domande in merito al regime a cui questi dati sono soggetti: di chi sono questi dati? Del soggetto che li fornisce (in questo caso l'amministrazione) o del soggetto che li detiene (in questo caso l'azienda che offre il servizio)? Quali implicazioni giuridiche derivano dal fatto che le pubbliche amministrazioni potenzialmente potrebbero non accedere per un determinato periodo ai dati in loro possesso? Che tipologia di contratto o accordo è possibile o auspicabile che ci sia tra le amministrazioni e i gestori di questi servizi di archiviazione? Quale politica di archiviazione dei dati, infine, è più adeguata per garantire i dati pubblici: il ricorso al cloud computing o le politiche di modernizzazione tecnologica anche avendo a mente la repentino ammodernamento delle nuove tecnologie?

Oltre gli aspetti meramente tecnologici e infrastrutturali, è interessante anche analizzare l'adeguatezza dell'attuale disciplina delle banche dati, soprattutto quella delle c.d. banche dati nazionali. La disciplina delle banche dati di rilievo pubblico (qui si tralasceranno tutti gli aspetti inerenti le banche dati private) diventa centrale per cogliere l'adeguatezza dell'ordinamento rispetto al fenomeno dei Big data. Gli aspetti più rilevanti del fenomeno, come la tendenza culturale a datizzare la realtà e a produrre dati e informazioni su tutto quello che ci circonda, rendono le banche dati e la loro disciplina gli strumenti principali per governare e interpretare concretamente questi aspetti. La disciplina delle banche dati nazionali, può essere considerato un tentativo di sistematizzare e razionalizzare la parte più consistente del patrimonio informativo pubblico con il fine di agevolare l'utilizzo dei dati e delle informazioni che sono contenute all'interno delle banche dati nazionali da parte di tutte le amministrazioni.

Qualsiasi valutazione sull'adeguatezza della disciplina delle banche dati permettendo alle amministrazioni di produrre e archiviare agevolmente dati e informazioni, è fortemente legata anche alle scelte di allocazione della funzione informativa all'interno del sistema amministrativo italiano (a livello centrale, regionale o locale). Utile infatti sarà, nel capitolo successivo, tentare di valutare l'adeguatezza per questi fini del regime delle banche dati nel nostro ordinamento.

L'impatto sull'adeguamento tecnologico che abbiamo descritto per le tecnologie di conservazione e produzione dei dati, vale in gran parte anche per le tecnologie che elaborano i Big data. Nei paragrafi iniziali abbiamo potuto vedere come l'elaborazione dei Big data necessitano di tecnologie che siano capaci di elaborare dati molto differenti tra di loro, alcune volte appartenenti a basi di dati differenti, che lavorano attraverso correlazioni e algoritmi matematici, quindi le amministrazioni si devono dotare di tecnologie capaci di elaborare i Big data.

Un'altra questione su cui il fenomeno dei Big data influisce, in maniera molto più evidente rispetto al regime dei dati e delle banche dati, e sulla quale produce certamente degli effetti giuridici da analizzare è il rapporto tra Big data e il regime degli Open data, sia riguardo la disciplina del Codice dell'amministrazione digitale inerente l'apertura delle banche dati ai cittadini, sia quella presente all'interno del decreto legislativo n. 33/2013. La possibilità di consultare ed estrapolare i dati e le informazioni contenute all'interno delle banche dati o all'interno dei siti istituzionali, non solo è fortemente legato con la tematica della qualità dei dati e di quali dati e quali informazioni possono essere resi pubblici e disponibili, ma anche legata alla realizzazione concreta di una nuova cittadinanza amministrativa. Chiarire, quindi, il rapporto tra Big data e Open data, è utile non solo per chiarire una differenza concettuale molto rilevante, ma anche per completare un discorso più generale sulla costruzione di una nuova cittadinanza amministrativa.

Ripensare il regime dei dati e delle banche dati e soprattutto costruirlo in via del tutto autonoma rispetto all'azione amministrativa che di volta in volta l'amministrazione è chiamata a compiere, caratterizza la dimensione organizzativa della funzione conoscitiva delle amministrazioni pubbliche ed è proprio in quest'ottica che le riflessioni sul regime dei dati e delle banche dati deve essere fatto.

I Big data pongono all'amministrazione pubblica una serie di problemi anche sul piano organizzativo. Il primo è legato, appunto, alla funzione conoscitiva autonoma, la quale ha necessariamente bisogno di strutture organizzative che la mettano in atto. Costruire un percorso razionale di gestione dei dati e saperli utilizzare nell'azione amministrativa è innanzitutto un problema organizzativo. Certamente i Big data interrogano le amministrazioni sull'adeguatezza dei sistemi informativi, sulla loro capacità di gestire

davvero i dati in possesso delle amministrazioni, sul loro livello di tecnicità, sul coordinamento informativo e sull'interoperabilità delle infrastrutture informatiche, sulle dinamiche di coordinamento e di collaborazione tra la parte dell'organizzazione che si occupa più strettamente di gestione ed elaborazione dei Big data e la restante parte dell'organizzazione amministrativa. I Big data, quindi, interrogano le amministrazioni e il sistema amministrativo italiano sulla possibilità di ricostruire una loro nuova capacità conoscitiva e tecnica.

*5. Big data e Open data: differenze concettuali e nuovi equilibri tra trasparenza e privacy.*

Ragionando sull'impatto dei Big data sull'organizzazione amministrativa, inoltre, non si può ignorare la necessità di un adeguamento delle competenze dei funzionari delle pubbliche amministrazioni.

Questo problema si pone non soltanto per quelle amministrazioni o per determinati uffici che si occupano più direttamente di gestione dei dati, di attività conoscitiva o di informatica, come gli uffici e le amministrazioni che si occupano di statistiche, di conoscenza pubblica in generale o i sistemi informativi presenti nelle amministrazioni, ma è tema complessivo che riguarda tutti i funzionari pubblici al di là delle singole mansioni. Tutti i funzionari pubblici devono avere una certa padronanza dello strumento, che non significa necessariamente avere competenze matematiche o statistiche pronunciate, ma avere le basi per poterlo utilizzare concretamente e soprattutto averne interiorizzato le opportunità e i limiti. Utilizzare i Big data nelle amministrazioni pubbliche presuppone, innanzitutto, un approccio culturale differente verso la gestione dei dati e verso le informazioni. La datizzazione della realtà, l'elaborazione dei Big data con metodi statistici e matematici lontani dal classico approccio causale o senza ipotesi di partenza, l'estrema valorizzazione del dato che rende recessiva non solo la sua qualità, ma valorizza i suoi possibili utilizzi secondari costituiscono una sfida culturale rilevante per i dirigenti e per i funzionari pubblici.

quanto la disciplina dell'azione amministrativa sia adeguata ad interpretare le nuove tendenze prodotte dal fenomeno dei Big data. Le potenzialità conoscitive dei Big data, la sua capacità di portare nuove conoscenze sui fenomeni, le sue capacità predittive modificano profondamente l'attività conoscitiva delle amministrazioni pubbliche. L'istruttoria procedimentale, il luogo che la legge n. 241/1990 riserva all'accertamento dei fatti e alla valutazioni degli interessi pubblici e privati che sono in gioco all'interno di un determinato procedimento, vede modificare i suoi principi fondanti. I Big data

modificano, vedremo quanto in positivo o in negativo, il principio di informalità e il metodo inquisitorio che caratterizzano l'attività istruttoria nel nostro ordinamento e mettono in evidenza come l'attività conoscitiva delle pubbliche amministrazioni non possa essere più considerata come un'attività strumentale all'azione amministrativa, a costo di svilirne la sua effettività di funzione atta a permettere all'amministrazione di conoscere i bisogni e le necessità delle comunità di riferimento e sulle quali le proprie decisioni producono effetti rilevanti.

La stessa decisione amministrativa, proprio grazie alle potenzialità conoscitive che l'elaborazione dei Big data produce, assume dei caratteri differenti, di maggiore certezza, di maggiore completezza. Questo non può esimerci dal ritornare ad affrontare il rapporto, sempre complesso, tra verità materiale, conoscenza oggettiva dei fenomeni e decisione amministrativa.

Sarà utile anche tentare di comprendere se l'utilizzo dei Big data produca effetti solo sull'attività conoscitiva o se possa modificare, in determinate occasioni, profondamente l'azione amministrativa anche condizionando alcuni aspetti della discrezionalità amministrativa. Il concreto utilizzo dei Big data all'interno del procedimento amministrativo, l'intensità del loro utilizzo e l'esclusività dei risultati della loro elaborazione come fondamento della decisione amministrativa, può modificare la discrezionalità amministrativa, ma anche fare riflettere sulla trasformazione dell'attività conoscitiva delle amministrazioni in un vero e proprio potere conoscitivo, capace di comprimere i diritti dei cittadini. Nel secondo capitolo si cercherà di indagare, in conclusione come e se le garanzie procedurali si trasformano con l'utilizzo dei Big data, con una particolare attenzione alla disciplina dei diritti di partecipazione dei cittadini al procedimento.

Nel terzo capitolo, invece, indagheremo sull'adeguatezza della disciplina dell'organizzazione amministrativa e sulla sua capacità di interpretare le esigenze di governo dei *Big data* all'interno delle amministrazioni pubbliche. Il primo aspetto che verrà indagato è quello inerente la gestione del patrimonio informativo pubblico. Un patrimonio informativo che con i Big data diventa più complesso, più eterogeneo e che pone all'ordinamento la necessità di adeguare a questa nuova eterogeneità la garanzia di qualità dei dati pubblici e il regime dei dati in generale. La possibilità di utilizzare concretamente i Big data passa anche da un adeguamento dell'organizzazione degli uffici dell'amministrazione e delle competenze dei funzionari pubblici. Ci si interrogherà se utilizzare i Big data possa o debba essere considerata una competenza specialistica o una competenza diffusa in tutti gli uffici e in possesso di tutti i funzionari pubblici, così come si cercherà di riflettere se i sistemi informativi e gli uffici tecnici delle amministrazioni siano adatti a gestire i Big data. In questo capitolo, infine, si

cercherà di comprendere come il coordinamento amministrativo cambia con l'utilizzo dei Big data e come la funzione di informazione e la funzione conoscitiva debba essere distribuita tra i vari livelli di governo.

L'ultimo capitolo si cercherà di tirare le fila del discorso innanzitutto evidenziando quali sono le criticità che emergono dalla ricostruzione normativa delle disposizioni dell'attuale disciplina dell'azione e dell'organizzazione amministrativa che più di tutte vengono toccate dall'espansione del fenomeno dei Big data. La generale inadeguatezza della disciplina positiva pone l'ordinamento di fronte alla necessità di affrontare gli effetti del fenomeno dei Big data in maniera complessiva con l'obiettivo di costruire un vero Big data government, cioè sapere governare i big data per poter governare con i Big data. Il capitolo, infatti, sarà un capitolo propositivo in cui si cercherà da un lato di limitare le declinazioni negative del fenomeno, ma dall'altro si cercherà di proporre delle soluzioni che non sviliscano le numerose potenzialità che derivano dall'utilizzo dei Big data. Il potere conoscitivo che l'elaborazione e l'utilizzo dei Big data nel procedimento amministrativo forma e consegna all'amministrazione deve essere tipizzato e regolato, riportandolo sotto l'influenza del principio di legalità, senza limitare la sua enorme portata conoscitiva, la quale avvicina l'amministrazione alla conoscenza oggettiva e completa dei fenomeni che è chiamata a governare, alla decisione ottimale sui problemi che è chiamata a risolvere e sui bisogni che è chiamata a garantire. La discrezionalità amministrativa non deve appiattirsi eccessivamente sui risultati delle elaborazioni dei Big data, ma deve anche potere fare tesoro dei risultati che provengono da quelle elaborazioni. Il regime dei dati, della loro gestione e della loro utilizzazione e riutilizzo da parte dei cittadini deve tenere conto delle necessità legate al rispetto della qualità dei dati e del possibile affidamento che i cittadini ripongono sulla provenienza pubblica dei dati, della privacy dei cittadini e delle nuovi potenziali rischi per la riservatezza dei dati personali che l'utilizzo dei Big data si porta dietro, ma deve anche garantire la massima diffusione dei dati e delle informazioni. Politiche di Open data sui Big data diventano in un contesto di forte accentramento informativo le uniche politiche pubbliche capaci di ristabilire gli equilibri democratici che le asimmetrie informative create dai Big data tendono ad incrinare. La diffusione e il riutilizzo dei Big data pubblici crea le condizioni per creare una nuova cittadinanza consapevole e capace di raccogliere e formare le informazioni attraverso le quali controllare l'operato delle amministrazioni e di partecipare consapevolmente ed attivamente alla decisione pubblica, garantendo anche un ritorno informativo per le amministrazioni, le quali magari non possedevano determinate informazioni.

## *6. Una nuova cittadinanza amministrativa con i Big data.*

Le fila di un ragionamento tuo, più innovativo: le trasformazioni del sistema amministrativo alla luce dei Big data, come impatterà, come il sistema lo regge, lo scenario. Qui devi agganciare il tema non tanto ai Big data, ma a tutta la dottrina solita. L'ingresso dei Big data nella ponderazione istruttoria. Posso utilizzare i Big data come elemento che contribuisce alla decisione? Discrezionalità e potere conoscitivo. Il principio inquisitorio ci impone di utilizzare questo nuovo strumento conoscitivo che permette di interpretare meglio la realtà e di utilizzarlo in quanto sia utilizzabile. Oppure dobbiamo attivare il procedimento laddove i Big data ci mostrino che ci sono fenomeni potenziali che la missione affidata all'amministrazione è quella di curare un certo interesse. Big data for better governance: buon andamento imparzialità come si pongono su questo tema? Letteratura dottrina, ragionamento sui dati e la base da cui partire per un ulteriore sviluppo che cambia in termini quantitativi o un'ulteriore passo avanti o parallela con un ulteriore passo avanti e l'approccio Open data e Big data: uso secondario dei dati, importanza del dato al di là dell'istruttoria e del singolo utilizzo, la qualità dei dati e Big data (diversa qualità, non importa, importa meno, non ho bisogno di irrigimentare i dati, ma ho bisogno di tutti i dati possibili).

Privacy e Big data; Open data e Big data; Big data per la regolazione; Big data e disciplina di settore; Primo approccio è governarle lei stessa? Regolarne l'utilizzo (forse nel secondo)?

Problemi: l'amministrazione è in grado di garantire i tempi che sono quelli che caratterizzano i Big data, quale competenze al suo interno? La pubblica amministrazione di domani potrebbe essere uno dei grandi fruitori o utilizzatori di Big data. Al fianco di Google o di Facebook domani ci sarà lo Stato italiano (o alcune sue parti specializzate)? I Big data sono la manifestazione più plastica del passaggio ad internet come nuova dimensione dei fenomeni sociali (come Terra e Mare di Carl Schmitt), le prime sentenze sul mare sono adesso (quella su Google per esempio). I fenomeni sociali stanno là dentro e quello ci consente una comprensione della realtà diversa, maggiore, cambia la qualità dell'approccio ai fenomeni e noi non siamo in grado di farlo per alcune cose potremmo farne un uso intelligente (sanità). Alcune cose che hai già visto le puoi recuperare anche quando parli di pubblica amministrazione per dire. Conclusioni articolate o ampie o un quinto sui problemi aperti (o evocativo).

Big data e tempi perché noi ragioniamo. L'amministrazione e i tempi di azione costruiti sul causa-effetto, spesso, ma costruiti su processi che richiedono tempo. Il discorso

predittivo cambia il concetto di tempo, anticipa. Oppure la possibilità di esaminare il fenomeno nel corso del suo svolgimento per provare a governarne lo svolgimento e non a indirizzarlo e a sanzionarlo. Attraverso i Big data l'azione non è più prima o dopo (ti indirizzo e ti controllo). In prospettiva l'azione è un'azione “durante”, pone un problema di ripensamento del ruolo dell'amministrazione e delle modalità di procedere. Il quarto capitolo può essere un capitolo di ragionamento, ma anche un capitolo in cui noi diciamo attenzione una serie di categorie sono in crisi: tempo, procedimento, indirizzo e controllo, discrezionalità.

Questo discorso come ultimo capitolo conclusivo. Nell'ultimo metti i problemi, nel penultimo tu sviluppi un ragionamento tuo, innovativo, in cui analizzi come cambia l'attività conoscitiva con i dati e con i Big data c'è un'ulteriore salto

Tutti gli studi analizzati partono dalla premessa che l'elaborazione dei Big data – cioè la capacità di conoscere meglio la realtà, la capacità di poter prevedere tendenze e fenomeni anche in un'ottica differente e imprevedibile rispetto alle classiche chiavi di lettura - possa migliorare notevolmente il processo decisionale pubblico, rendendolo maggiormente efficace, democratico, ma anche pervasivo. In questo senso vanno sia studi meno recenti - come quello di David Bollier “*The promise and perile of Big data*” per l'Aspen Institute (2010) o quello dell'Economist Intelligence Unit “*Big Data and the Democratisation of Decisions*” (2012) – sia studi più recenti – come lo studio compiuto da alcuni studiosi per la Commissione europea “*Policy Making 2.0: Unleashing the power of big data for public governance*” (2014).

Lo scopo del progetto dell'unità di ricerca di Perugia è quello di riflettere sull'implicazione delle nuove modalità di sfruttamento dei dati, riconducibili al paradigma “big data”, nell'ottica della trasformazione e del miglioramento della governance pubblica. Si tratta di una questione che è vista come centrale nei documenti che definiscono le coordinate di fondo del nuovo “big data approach” (Us, Ocse), dove sia con riferimento ad alcuni focus settoriali (in particolare, i servizi sanitari, i servizi di trasporti, l'accertamento fiscale), sia con riferimento a specifiche dimensioni territoriali (le “smart cities”), sia ponendo attenzione a specifiche funzioni (come la statistica ufficiale), la questione dello sfruttamento dei dati per il miglioramento dell'efficienza e dell'efficacia dell'azione pubblica è posta con evidenza. Per confrontarsi con tale questione è necessario porre attenzione al regime dei dati pubblici ed alle condizioni di conservazione, fruizione, sfruttamento e valorizzazione del patrimonio informativo pubblico, sia direttamente da parte delle pubbliche amministrazioni, sia indirettamente attraverso politiche di promozione di un ruolo attivo dei privati nell'ottica di una innovazione sociale attraverso l'apertura di basi di dati (“open innovation”), e rispetto alle quali è necessario riflettere sul ruolo dei soggetti deputati a regolare il regime di

questo “nuovo” trattamento, sia allorché ci si trovi di fronti a dati personali, o suscettibili di tornare tali (la “de-anonimizzazione” grazie alle tecniche di big data), sia allorché ci si ponga nell'ottica della qualità e del coordinamento dei dati in vista di un loro più ampio e agevole sfruttamento (una questione, quest'ultima, che chiama in particolare in causa il ruolo di Agid e le strategie di valorizzazione a fini conoscitivi dei dati). Rispetto alle coordinate tradizionali dell'attività conoscitiva pubblica, anche ove intesa come tendenza alla autonomizzazione della funzione conoscitiva rispetto al suo ruolo di norma “servente” (e a perdere) nei singoli procedimenti, l'epoca dei big data segnala uno scarto, che va oltre la recente evoluzione che presuppone un miglioramento della capacità conoscitiva pubblica connessa agli utilizzi secondari dei dati e all'integrazione delle basi di dati. La prospettiva di un utilizzo predittivo e valutativo dei big data produce, a ben vedere, cambiamenti quantitativi e qualitativi nell'ottica della amministrative investigation, ponendo problemi rilevanti alla luce dei principi che tradizionalmente reggono l'attività conoscitiva pubblica: fino a che punto una decisione può essere fondata su elementi “indiziari”? Quali sono le esperienze, le “best practices” e i problemi aperti nei contesti, anche esteri, che per primi hanno sperimentato il ricorso a tecniche di big data government a supporto di decisioni e politiche pubbliche? Quali i limiti a fronte di altre regole e principi sul trattamento dei dati e delle informazioni? Quali differenze rispetto all'azione ed allo sfruttamento dei dati da parte di soggetti privati? E, de jure condendo, quale strategia di regolazione del big data government?

I risultati che si attendono da un'analisi di questo genere sono essenzialmente due. Innanzitutto ci si attende l'emersione degli elementi problematici e le nuove necessità che il fenomeno pone alle pubbliche amministrazioni. Le norme attuali sul sistema informativo, sulle banche dati, sull'istruttoria procedimentale, sulla privacy in relazione alle amministrazioni pubbliche non sono del tutto adeguate per affrontare le sfide del fenomeno dei Big data. In secondo luogo ci si pone l'obiettivo di sistematizzare gli elementi problematici e provare a riflettere su un sistema di regole capace di agevolare una nuova cultura di governo dei dati nelle pubbliche amministrazioni; un nuovo modo di amministrare i Big data (per amministrare con i Big Data); una disciplina dell'azione e dell'organizzazione amministrativa che consenta di sfruttare le potenzialità dei Big data e di contenerne i limiti; nuove competenze da integrare nell'amministrazione; una nuova cittadinanza attiva che sappia sfruttare per prima Big data in funzione di controllo democratico, di tutela personale e di partecipazione alla decisione. L'impatto della ricerca è quello di fornire un quadro chiaro sulle potenzialità del Big data government in Italia. All'estero (Australia *The Australian Public Service Big Data Strategy: Improved Understanding through Enhanced Data-analytics Capability Strategy Report*. 2013; USA con *Big data: seizing opportunities, preserving values* 2014) si stanno interrogando concretamente su come utilizzare i Big data per migliorare il processo

decisionale in molti settori (in particolare difesa, sanità, sicurezza), molte amministrazioni pubbliche hanno incominciato a delineare quali sono le potenzialità del Big data government: rendere i decisori politici più consapevoli della realtà in cui governano (anche con nuovi punti di vista); le amministrazioni pubbliche tecnicamente più capaci, potenzialmente più imparziali (perché in possesso di più elementi conoscitivi) e più efficienti; i cittadini più consapevoli del loro ruolo nell'era dei Big data, in cui non si limitano a partecipare alla decisione pubblica in maniera più consapevole, ma utilizzano ed elaborano i dati per controbilanciare il nuovo potere conoscitivo pubblico, al fine di attuare un migliore controllo democratico dell'operato della pubblica amministrazione.

## BIBLIOGRAFIA

- AA.VV., *Counting every moment*, in *The Economist*, 3 marzo 2012.
- AA.VV., *The Big Data Conundrum: How to Define It?*, MIT Technology Review, 3 ottobre 2013.
- ALLEGRETTI U., *Amministrazione pubblica e Costituzione*, Padova, 1991;
- ALLEVA G., *Conservazione e accessibilità dei microdati prodotti dall'ISTAT e cenni all'ambito del SISTAN*, Audizione del Presidente dell'ISTAT presso la Commissione per la garanzia della qualità dell'informazione statistica, 12 febbraio 2016;
- ANDERSON C., *The Petabyte Age*, in *Wired*, giugno 2008.
- ARENA G., *L'utente-sovrano*, in *Studi in onore di Feliciano Benvenuti*, Modena, 1996;
- AZZARITI G., *Lex digitalis e democrazia*, in M. NISTICÒ – P. PASSAGLIA (A CURA DI), *Internet e Costituzione*, Torino, 2014.
- BACHELET V., *Costituzione e amministrazione. Scritti giuridici*, Roma, 1992;
- BARTOLE S., *Interpretazioni e trasformazioni della Costituzione repubblicana*, Bologna, 2004;
- BASILAVECCHIA M., *Accertamento (Dir. Trib.)*, in CASSESE S. (a cura di), *Dizionario di diritto pubblico*, Milano, 2006, p. 45 e ss..
- BASILAVECCHIA M., *Accertamento (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, Milano, 2006.
- BASILAVECCHIA M., *Accertamento e studi di settore: soluzione finale*, in *Riv. Giur. Trib.*, 2010, nota di Cass., SS.UU, 18.12.2009, n. 26638.
- BASILAVECCHIA M., *Privacy e accertamento sintetico: primi segnali di riequilibrio*, in *Corr. Trib.*, 2014.
- BASILAVECCHIA M., *Problemi veri e falsi del nuovo redditometro*, in *Corr. Trib.*, 2013, fasc. 27, nota di Comm. Trib. Prov. - Reggio Emilia, sez. II, 18 aprile 2013, n. 74.
- BASTIONI FERRARA F., *I principi della riforma tributaria: accertamento sintetico e redditometro*, in *Riv. Dir. Trib.*, 1994, I.
- BAUMAN Z., LYON D., *Sesto potere. La sorveglianza nella modernità liquida*, trad. it. M. Cupellaro, Bari, 2014.

- BEGHIN M., *Il redditometro e gli altri accertamenti per "standard" nelle maglie della presunzione semplice*, in *Corr. Trib.*, 2013, nota Cass., sez. trib., 20.12.2012, n. 23554.
- BEGHIN M., *Profili sistematici e questioni aperte in tema di accertamento «sintetico» e «sintetico-redditometrico»*, in *Riv. Dir. Trib.*, 2010, I.
- BENVENUTI F., *Il nuovo cittadino. Tra libertà garantita e libertà attiva*, Venezia, 1994;
- BEYER M. A. - LANEY D., *The importance of big data: A definition*, Stamford, Gartner Retrieved, 21 June 2012;
- BINNEY W., *Sworn Declaration in the Case of Jewel vs. NSA*, 2 luglio 2012.
- BOER P., *L'attività di accertamento, vigilanza ed ispezione degli Enti previdenziali*, in *Mass. Giur. Lav.*, 1996, p. 679.
- BOGNI M. – DEFANT A., *Big data: diritti IP e problemi della privacy*, in *Dir. Ind.*, fasc. 2, 2015;
- BOLLIER D., *The promise and peril of Big data*, Washington, *The Aspen Institute*, 2010;
- BOYD D. – CRAWFORD K., *Six Provocations for Big Data*, presentato al convegno dell'Oxford Internet Institute "A Decade in Internet Time: Symposium on the Dynamics of Internet and Society", 21 settembre 2011;
- CARETTI P. – MORISI M., *Parlamento e politiche pubbliche*, in *Quad. cost.*, fasc. 3, 2001.
- CAVALLO-PERIN R., *La configurazione della cittadinanza amministrativa*, in *Dir. amm.*, 2004, 202 ss.
- CERULLI IRELLI V. (a cura di), *La disciplina generale dell'azione amministrativa*, Napoli, 2006;
- CHITI M. P., PALMA G. (a cura di), *I principi generali dell'azione amministrativa*, Napoli, 2006;
- CIANELLA P – CIANELLA S., *Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro*, in *D. disc. Priv., sez. comm.*, VIII, Torino, 1992.
- CIANELLA S. – ZOPPOLI L., *Istituto Nazionale di Previdenza Sociale*, in *D. disc. Priv., sez. comm.*, VIII, Torino, Utet, 1992.
- CINELLI M., *Diritto alla previdenza sociale*, Torino, 1998;
- CINELLI M., *Sicurezza sociale*, in *Enc. Dir.*, XLII, Milano, 1990, p.4;

Circolare dell'Agenzia delle Entrate, 11 marzo 2014, n. 6/E che recepisce le osservazioni del Garante e integra la Circolare dell'Agenzia delle Entrate, 31 luglio 2013, n. 24/E di commento generale alla disciplina del redditometro.

COGNETTI S., *Principio di proporzionalità. Profili di teoria generale e di analisi sistematica*, Torino, 2011;

COLOMBO D., *Con i Big data l'INPS stima il ricorso alla CIG*, Sole 24 ore del 17 marzo 2013 e consultabile on-line su <http://24o.it/5ocJ9>

COM (2014) 442, Comunicazione della Commissione europea al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, *Verso una florida economia basata sui dati*;

COM (2014) 442, *Comunicazione della Commissione europea al Parlamento europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni, Verso una florida economia basata sui dati.*

*Comunicazione della Commissione sulla definizione di mercato rilevante ai fini dell'applicazione del diritto comunitario in materia di concorrenza* [Gazzetta ufficiale C 372 del 9.12.1997].

CONTE D., *Aspetti problematici del redditometro e di altri consimili metodi di accertamento come strumenti di contrasto all'evasione*, in *Riv dir. Trib.*, 2008, I;

CONTI G. L., *La governance dell'Internet: dalla costituzione della rete alla costituzione nella rete*, in M. NISTICÒ – P. PASSAGLIA (A CURA DI), *Internet e Costituzione*, Torino, 2014.

CONTRINO A. – MARCHESELLI A., *Il «redditometro 2.0» tra esigenze di privacy, efficienza dell'accertamento e tutela del contribuente*, in *Dir. E Prat. Trib.*, fasc. 4, 2014.

CORRADO L. R., *Accertamenti standardizzati e motivazione dell'avviso di accertamento: l'atto è illegittimo in difetto di una adeguata replica alle deduzioni fornite dal contribuente in sede di contraddittorio endoprocedimentale*, in *Dir. Prat. Trib.*, 2008, II, Cass., Sez. trib., 22 febbraio 2008, n. 4624.

CORRADO L. R., *Il contraddittorio endoprocedimentale condiziona l'efficacia probatoria degli studi di settore*, in *Corr. Trib.*, 39, 2015.

CROSBY A. W., *The Measure of reality*, Cambridge University Press, Cambridge, 1996, tr. it., *La misura della realtà*, Dedalo, Bari, 1998.

CUOCOLO L., *I rapporti stato-regioni in materia farmaceutica, tra diritti costituzionali e contenimento della spesa*, in *Quad. reg.*, 2008, fasc. 1.

DAMIANI M., *Nessuna prevalenza tra studi di settore e criterio di tassazione per cassa*, in *Corr. Trib.*, n. 38/2010, nota di Cass., Sez. trib., 9 luglio 2010, n. 16235

DAVENPORT T.H. – BARTH P. – BEAN R., *How Big data is different*, in *Sloan review*, 30 luglio 2012.

DE MAURO A. – GRECO M. – GRIMALDI M., *What is big data? A consensual definition and a review of key research topics*, *AIP Conference Proceedings*.

DE PADOVA C., *Big data, la nuova frontiera dell'IT*, in *Riv. Inf. Mal. Prof.*, 2013, fasc. 1-2.

DE SIERVO U., *Ma chi fa i decreti legislativi?*, in COCOZZA V. – STAIANO S. (A CURA DI), *I rapporti tra Parlamento e Governo attraverso le fonti del diritto*, I, Torino, 2001.

DEL MONDO G., *Big Data e Analytics: modelli, analisi e previsioni, per valorizzare l'enorme patrimonio informativo pubblico*, in [https://www.inps.it/docallegati/InpsComunica/ForumPA2013/Documents/Del%20Mondo-ForumPA\\_2013\\_definitiva.pdf](https://www.inps.it/docallegati/InpsComunica/ForumPA2013/Documents/Del%20Mondo-ForumPA_2013_definitiva.pdf).

DELLA VALLE E., *Note minime in tema di accertamenti standardizzati*, in *Rass. Trib.*, 2014, fasc. 4.

DEMURO G., *La leale collaborazione nella differenziazione della spesa farmaceutica*, in *Le Regioni*, 2012, fasc. 1-2;

Department of Education, *Protecting Student Privacy While Using Online Educational Services: Requirement and Best Practices*, February, 2014.

Department of Homeland Security, *Privacy Impact Assessment for the Cerberus Pilot*, November, 2013.

Department of Homeland Security, *Privacy Impact Assessment for the Neptune Pilot*, September 2013;

DI SABATO D., *La congruità del contribuente e la coerenza del sistema*, in *Riv. Dir. Dell'impr.*, 2014, fasc. 2;

Dipartimento di Finanza, *The Australian Public Service Big Data Strategy: Improved Understanding through Enhanced Data-analytics Capability Strategy Report*, 2013;

DIXON P. – GELLMAN R., *The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future*, *World Privacy Forum*, April 2014.

- DOGLIANI M. (A CURA DI), *Il libro delle leggi strappato e la sua manutenzione*, Torino, 2012;
- DONVITO D., *Big data e Statistica Ufficiale: le strategie ISTAT*, in <http://docplayer.it/958327-Big-data-e-statistica-ufficiale-la-strategia-istat.html>.
- DUGAS A. F. ET AL., *Google Flu Trend: Correlation with Emergency Department Influenza Rates and Crowding Metrics*, CID Advanced Access, 8 gennaio 2012.
- DUGGAN M. – LEVITT S.D., *Winning Isn't Everything: Corruption in Sumo Wrestling*, in *American Economic Review*, 92 (2002).
- DUHIGG C., *The power of habit: why we do what we do in life and business*, Random House, New York, 2012, tr. it., *La dittatura delle abitudini*, Corbaccio, Milano, 2012.
- E. MOROZOV, *L'ingenuità della rete. Il lato oscuro della libertà di internet*, Torino, 2011;
- ETZIONI O., KNOBLOCK C.A., TUCHINDA R. - YATES A., *To Buy or Not to Buy: Mining Airfare Data to Minimize Ticket Purchase Price*, SIGKDD '03, 24-25 agosto 2003.
- European Data Protection supervisor, *Privacy and competitiveness in the age of big data: The interplay between data protection, competition law and consumer protection in the Digital Economy*, March 2014.;
- Executive Office of the President - Office of Science and Technology Policy, *Fact Sheet: Big Data Across the Federal Government*, Washington, marzo 2012.
- Executive Office of the President of the U.S.A., *Big data and differential pricing*, Washington, Febbraio 2015;
- Executive Office of the President of the U.S.A., *Big data: seizing opportunities, preserving values*, Washington, Maggio 2014;
- Executive Office of the President of the U.S.A., *President's Council of Advisors on Science and Technology, Big Data and Privacy: a technological perspective*, Maggio 2014.
- FALSITTA G., *Manuale di diritto tributario. Parte generale*, Padova, 2012;
- Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Business and Consumers*, 2012.
- FERRARI G. – FERRARI G., *Infortuni sul lavoro e malattie professionali*, Padova, IV, 2002.

- FROSINI T.E., *Internet come ordinamento giuridico*, in NISTICÒ M. – PASSAGLIA P. (A CURA DI), *Internet e Costituzione*, Torino, 2014.
- FURNAS A., *Homeland Security's «Pre-Crime» Screening Will Never Work*, in *The Atlantic Online*, 17 aprile 2012.
- GAFFURI G., *I metodi di accertamento fiscale e le regole costituzionali*, in *Studi in onore di Uckmar*, Padova, 1997.
- GALETTA D. U., *Proporzionalità e sindacato giurisdizionale nel diritto amministrativo*, Milano, 1998;
- GALLO F., *Accertamento e garanzie del contribuente*, in *Dir. Prat. Trib.*, 1989, I;
- GALLO F., *Discrezionalità dell'accertamento tributario e sindacabilità delle scelte dell'ufficio*, *Riv. Dir. Fin.*, 1992;
- GANZ J. – REINSEL D., *Big data, bigger digital shadows and biggest growth in the Far East*, 1 dicembre 2012.
- Garante per la protezione dei dati personali, *Parere sul PSN 2014-2016 aggiornamento 2015-2016*, 18 settembre 2014 n. 411.
- Garante per la protezione dei dati personali, *Parere sullo schema di Programma statistico nazionale (Psn) 2014-2016 - Aggiornamento 2016*, del 29 ottobre 2015, n. 566.
- GARLAND D., *The Culture of Control*, University of Chicago Press, Chicago, 2001.
- GERON T., *Twitter's dick costolo: Twitter mobile and revenue beats desktop on same days*, in *Forbes*, 6 giugno 2012.
- GIGLIONI F., *La sfida dell'innovazione sulla regolazione pubblica. il caso delle "smart grid"*, in *Munus*, 2013, fasc. 3.
- GIGLIONI F., *Le soluzioni istituzionali alla qualità dell'informazione statistica*, in *Dir. Pubbl.*, 2005.
- GINSBURG J. ET AL., *Detecting Influenza Epidemics Using Search Engine Query Data*, *Nature*, 457, 2009.
- GIORDANO S., *Potenziati e irreparabili danni alla libertà personale con il "nuovo redditometro"*, in *GT – Riv. Giur. Trib.*, 2013, fasc. 4, nota di Trib. Napoli, Sez. di Pozzuoli, Ord. 21 febbraio 2013.
- GIRELLI G., *Gli studi di settore quale strumento "multifunzionale" tra dichiarazione, accertamento e processo tributario*, in *Riv. Dir. Trib.*, 2012, fasc. 7-8;

- GIUBBONI S. – CINELLI M., *Il diritto della sicurezza sociale in trasformazione*, Torino, 2005.
- GONZALEZ J., *Whistleblower: The NSA Is Lying – U.S. Government Has Copies of Most of Your Emails*, in *Democracy Now*, 20 aprile 2012;
- GRASSETTI C., *Statistica per la pubblica amministrazione*, Padova, 2008.
- GUERRA M. P., *Statistica pubblica e funzione conoscitiva*, in MERLONI F. (a cura di), *L'informazione delle pubbliche amministrazioni*, Rimini, 2002;
- GUGLIELMETTI G., *La tutela della banche dati con diritto sui generis nella direttiva 96/9/CE*, in *Contr. e impr. Europa*, 1997.
- HAN H. – YONGGANG W. – TAT-SENG C. – XUELONG L., *Towards scalable systems for big data analytics: a technology tutorial*, *IEEE Access*, 2014.
- HARCOURT B. E., *Against Prediction: Profiling, policing and punishing in an actuarial age*, University of Chicago Press, Chicago, 2006.
- HAVELY A. - NORVING P. – PEREIRA F., *The Unreasonable effectiveness of data*, in *IEE Intelligence System*, marzo-aprile 2009.
- HILBERT M. – LOPEZ P., *The world's technological capacity to communicate, store and compute information?*, in *International Journal of Communication*, 2012.
- House of Commons, Science and Tecnology Committee, *The Big data dilemma*, Fourth Report of session 2016-16.
- INGLEHART R. – KLINGEMANN H.D., *Genes, Culture and Happiness*, MIT Press, Cambridge, 2000.
- ISTAT, *Linee strategiche del censimento permanente della popolazione e delle abitazioni, metodi, tecniche e organizzazione*, 2014.
- KANHEMAN D., *Thinking Fast and Slow*, Farrar, Straus and Giroux, New York 2011, pp. 74-75, tr. it., *Pensieri lenti e veloci*, Mondadori, Milano, 2012;
- KRANZBERG M., *Technology and History: Kranzberg's Laws*, in *Technology and Culture*, 1986.
- LA ROSA S., *Principi di diritto tributario*, Torino, 2004.
- LA SPINA A., *La protezione sociale*, in S. CASSESE (a cura di ), *Trattato di diritto amministrativo*.
- LAFAVE W., *Search and Seizure: A Treatise On The Fourth Amendment*, §§ 1.1–1.2 West Publishing, 5th ed. 2011.

- LAMETTI D., *Cloud computing: verso il terzo Enclosures Movement?*, in *Riv. critica dir. Priv.*, 3, 2012.
- LANEY D., *3d data management: Controlling data volume, velocity and variety*, Gartner, Stamford, CT, USA, White Paper, 2001.
- LEVI F., *L'attività conoscitiva della pubblica amministrazione*, Torino, 1967.
- LOVECCHIO L., *Spese per incrementi patrimoniali imputabili secondo il principio della quota risparmio anche nel "nuovo redditometro"*, in *Riv. giur. Trib.*, 2013, nota di Cass., sez. VI, ord. 6.2.2013, n. 2806
- LOWREY A., *Economists's Programs are beating U.S. at Tracking Inflation*, in *Washington Post*, 25 dicembre 2010.
- LUPI R., *Metodi induttivi e presunzioni nell'accertamento tributario*, Milano, 1988.
- LUTHY C., *Guest Perspective: IBM Working with PG&E to Maximize the EV Potential*, in *PGE Currents*, 13 aprile 2012.
- LYON D., *Massima Sicurezza. Sorveglianza e "guerra al terrorismo"*, trad. it. E. Greblo, Milano, 2005, pp. 163.164.
- MANTELERO A. – VACIAGO G., *Data protection in a Big data society. Ideas for a future regulation*, *Digital investigation*, November 2015.
- MANTELERO A., *Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo*, in *Il Diritto dell'informazione e dell'informatica*, 2012, fasc. 1.
- MANTELERO A., *Il futuro regolamento EU sui dati personali e la valenza "politica" del caso Google: ricordare e dimenticare nella digital economy*, in *Il Diritto dell'informazione e dell'informatica*, 2014.
- MANTELERO A., *The future of consumer data protection in the E.U. Rethinking the "notice and consent" paradigm in the new era of predictive analytics*, *Computer Law & Security Report*, November 2014.
- MANZONI I. – VANZ G., *Il diritto tributario*, Torino, 2008;
- MANZONI I., *Potere di accertamento e tutela del contribuente*, Milano, 1993.
- MARCELLI F., *Internet fra canale di partecipazione politica e strumenti di controllo. Profilo di diritto internazionale*, in MARCELLI F., MARSOCCI P., PIETRANGELO M. (A CURA DI), *La rete internet come spazio di partecipazione politica. Una prospettiva giuridica*, Napoli, 2015.

- MARCHESELLI A., *Accertamenti tributari e difesa del contribuente. Poteri e diritti nelle procedure fiscali.*, Milano, 2010.
- MARCHESELLI A., *Il “giusto precedimento” tributario. Principi e discipline*, Padova, 2012.
- MARCHESELLI A., *La Sezione Unite sulla natura presuntiva degli studi di settore*, in *Corr. Trib.*, 2010, nota di Cass., SS.UU., 18.12.2009, n. 26635
- MARCHESELLI A., *Le presunzioni nel diritto tributario: dalle stime agli studi di settore*, Torino, 2008.
- MARCHESELLI A., *Redditometro e diritti fondamentali: da Garante e giurisprudenza estera un «assist» ai giudici tributari italiani*, in *Corr. Trib.*, 2014.
- MARINI G., *Diritto alla privacy. Commento al Codice in materia di protezione dei dati personali*, in BARBA A. – PAGLIANTINI (A CURA DI), *Delle persone*, in GABRIELLI E. (DIRETTO DA), *Commentario del codice*, UTET, 2013.
- MASSARI G., *La sicurezza dei farmaci in italia, tra vincoli sopranazionali e riparto interno di competenze*, in *Ist. Fed.*, 2015, fasc. 1;
- MASSERA A., *I principi generali dell'azione amministrativa tra ordinamento nazionale e ordinamento comunitario*, in *Dir. amm.*, 2005;
- MATTARELLA B. G., *La trappola delle leggi. Molte, oscure e complicate*, Bologna, 2011;
- MAYER SCHONBERGER V. – CUKIER K., *Big data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, trad. it. R. Merlini, Garzanti, Milano, 2013.
- MAZZANTINI G., *Il decreto “spending review”*, in *Gior. dir. amm.*, 2012, fasc. 12.
- MCGREGOR C. - CATLEY C. - JAMES A. – PADBURY J., *Next Generation Neonatal Health Informatics with Artemics*, in European Federation for Medical Informatics, MOEN A. ET ALTRI (A CURA DI), *User Centred Networked Health Care*, IOS Press, Lansdale, 2011.
- MCKINSEY, *Big data: the next frontier for innovation, competition and productivity*, 2011;
- MENNINI et al., *Disponibilità a pagare e innovazione: il caso dei farmaci anti-HCV nel Sistema Sanitario Italiano*, GRHTA, 2015, 2(2), p. 69.
- MERLONI F., *Istituzioni di diritto amministrativo*, Torino, 2016.

- MERUSI F., *Ragionevolezza e discrezionalità amministrativa*, Napoli 2011;
- MINDUS P., *Libertà politica e nuove tecnologie*, in *Ragion pratica*, 44/2015.
- MISURACA G., MAREDDU F., OSIMO D., *Policy Making 2.0: Unleashing the power of big data for public governance*, in Springer / Heidelberg Dordrecht, 2014.
- MOSCHETTI F., *Avviso di accertamento tributario e garanzie per il cittadino*, in *Dir. Prat. Trib.*, 1983, I.
- NOCERA C., *Il contraddittorio nella recente prassi e giurisprudenza in materia di studi di settore*, in BORIA P. (a cura di), *Studi di settore e tutela del contribuente*, Milano, 2010.
- OCSE, *Data-driven innovation. Big data for growth and well-being*, 6 October 2015.
- ODLYZKO A., *Privacy, price discrimination and seeds of capitalism's destruction*, 28 ottobre 2015, Camera dei deputati organizzato dall'Intergruppo parlamentare sull'innovazione tecnologica;
- OHM P., *Broken promises of privacy: responding to the surprising failure of anonymization*, *UCLA Law Review*, 2010.
- PACINI F., *Dall'autorità della legge all'autorevolezza delle Assemblee*, in *Quad. cost.*, fasc. 1, 2015.
- PAOLANTONIO M., POLICE A., ZITO A. (a cura di), *La pubblica amministrazione e la sua azione*, Torino, 2005;
- PARDOLESI R., *Concorrenza sul mercato: produzione di farmaci, collocazione a carico dello stato ed esclusione di intese restrittive volte a promuovere l'inserimento nel mercato del più costoso*, in *Foro it.*, 2015, fasc. 1, nota di TAR Lazio - Roma sez. I 2 dicembre 2014, n. 1216.
- Parere del Garante per la protezione dei dati personali, *Redditometro: le garanzie dell'Autorità a seguito della verifica preliminare sul trattamento di dati personali effettuato dall'Agenzia delle entrate* del 21 novembre 2013.
- PASINI P. – PEREGO A. (A CURA DI), *Big data: nuove fonti di conoscenza aziendale e nuovi modelli di management*, Rapporto di ricerca per IBM, SDA Bocconi – School of Management, 2012.
- PASQUALE F., *The Black Box Society: The Secret Algorithm Behind Money and Information*, Harvard University Press, 2014;
- PASTORI G., *Diritti e servizi oltre la crisi dello Stato sociale*, in *Scritti Ottaviano*, Milano, 1993;

PCAST Report, *Big Data and Privacy*; Harvard Law, Petrie-Flom Center, *Online Symposium on the Law, Ethics & Science of Re-identification Demonstrations*, <http://blogs.law.harvard.edu/billofhealth/2013/05/13/online-symposium-on-the-law-ethics-science-of-re-identification-demonstrations/>

PERRONE L., *Discrezionalità amministrativa (Dir. Trib.)*, in in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, Milano, 2006.

PERRONE L., *Il redditometro verso accertamenti di massa (con obbligo del contraddittorio e dell'invito all'adesione)*, in *Rass. Trib.*, fasc. 4, 2011.

PIZZETTI F., *Datagate, Prism, caso Snowden: il mondo tra nuova grande guerra cibernetica e controllo globale*, in [www.federalismi.it](http://www.federalismi.it), fasc. 13, 2013.

PIZZORUSSO A., *La manutenzione del libro delle leggi ed altri studi sulla legislazione*, Torino, 1999;

President's Council of Advisors on Science & Technology, *Realizing the Full Potential Of Health Information Technology to Improve Health Care for Americans: The Path Forward*, The White House, December, 2010.

Presidenza del Consiglio dei ministri, *Strategia per la crescita digitale 2014-2020*, Roma, 3 marzo 2015, p. 111 ss.. Consultabile in [http://www.agid.gov.it/sites/default/files/documentazione/strat\\_crescita\\_digit\\_3marzo\\_0.pdf](http://www.agid.gov.it/sites/default/files/documentazione/strat_crescita_digit_3marzo_0.pdf).

PRIEST D. – ARKIN W., *A Hidden World, Growing Beyond Control*, in *The Washington Post*, 19 luglio 2010;

*Programma statistico nazionale (PSN) 2014-2016 – Aggiornamento 2016*.

RAGUCCI G., *Il contraddittorio nei procedimenti tributari*, Torino, 2009, p. 166 e ss..

RENNA M., SATTA F., *Studi sui principi del diritto amministrativo*, Milano, 2012;

RESTA G., *Identità personale e identità digitale*, in *Diritto dell'informazione e dell'informatica*, 2007.

REZZANI A., *Big data. Architettura, tecnologie e metodi per l'utilizzo di grandi basi di dati*, Maggioli, Bologna, 2013.

RODOTÀ S., *Il diritto di avere diritti*, Bari, 2012.

RODOTÀ S., *Of Machine and men*, in M. HILDEBRANDT, A. ROUVROY (A CURA DI), *Law, Human Agency and Autonomic Computing: The Philosophy of Law Meets the Philosophy of Technology*, New York, 2011, 179 ss.;

- RONDINI L.L., *Storia della statistica pubblica in Italia*, Milano, 2003;
- ROSENTHAL J., *Special Report: International Banking*, in *The Economist*, 19 maggio 2012.
- ROSSI G., *L'attualità di G.D. Romagnosi nell'eclissi dello statalismo. Considerazioni sul passato e sul futuro del diritto amministrativo*, in *Dir. pubbl.*, 2012.
- ROSSI P., *I metodi di accertamento*, in FANTOZZI A. (a cura di), *Diritto tributario*, IV ed., Torino, 2013.
- RUDIN C. ET AL., *21st-Century Data Miners Meet 19th Electrical Cables*, in "Computer", giugno 2011.
- RUDIN C. ET AL., *Machine Learning for the New York City Power Grid*, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34, n. 2 (2012).
- SALVINI L., *La partecipazione del privato all'accertamento*, Padova, 1990, p. 67 e ss.;
- SALVINI L., *Procedimento amministrativo (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, Milano, 2006, p. 4531;
- SANDULLI A., *La proporzionalità dell'azione amministrativa*, Padova, 1998;
- SANDULLI M.A. (a cura di), *L'azione amministrativa*, Milano, 2005;
- SANDULLI M.A., *Codice dell'azione amministrativa*, Milano 2011.
- SCAPPIERO M., *Esperti a confronto per orientare le scelte dell'ISTAT sui Big data*, in [http://www.istat.it/it/files/2013/06/big\\_data.pdf](http://www.istat.it/it/files/2013/06/big_data.pdf).
- SCHIAVOLIN R., *Poteri istruttori dell'amministrazione finanziaria*, in *Dig. Disc. Priv. Sez. Comm.*, XI, Torino, 1995.
- SCHWARZE J., *The Principle of Proportionality and the Principle of impartiality in European Administrative Law*, in *Riv. trim. dir. Pubbl.*, 2003;
- SCISM L.– MAREMONT M., *Inside Deloitte's Life-Insurance Assessment Technology*, in *Wall Street Journal*, 19 novembre 2010.
- SCISM L.– MAREMONT M., *Insurers Test Data Profiles to Identify Risky Clients*, in *Wall Street Journal*, 19 novembre 2010;
- SCIULLO G. (a cura di), *Le nuove regole dell'azione amministrativa*, Bologna, 2006;
- SPANDONARO F., *Le politiche farmaceutiche e i fondi per l'innovazione: come garantire la sostenibilità*, in [http://www.quotidianosanita.it/studi-e-analisi/articolo.php?articolo\\_id=20939](http://www.quotidianosanita.it/studi-e-analisi/articolo.php?articolo_id=20939).

- SUNSTEIN C. R., *Simpler: the future of government*, New York, Simon & Schuster, 2013.
- SWEENEY L., *Discrimination in Online Ad Delivery*, 2013, reperibile facilmente sul sito <http://dataprivacylab.org/projects/onlineads/1071-1.pdf>.
- SYLOS LABINI F., *Big data, complessità e metodo scientifico*, in *Aspenia*, n. 63, Aspen Institute Italia, dicembre 2013.
- THURM S., *Next Frontier in Credit Score: Predicting Personal Behavior*, *Wall Street Journal*, 27 ottobre 2011.
- TOMEI R. (a cura di), *La nuova disciplina dell'azione amministrativa*, Padova, 2005;
- TWDL, *Big Data Analytics*, 14 settembre 2011. Sul sito è possibile anche consultare una serie, anche recentissimi, di report sul tema.
- VANZ G., *Controlli amministrativi (Dir. Trib.)*, in S. CASSESE (a cura di), *Dizionario di diritto pubblico*, Milano, 2006.
- VASSALLO S., *Liberiamo la politica. Prima che sia troppo tardi*, Bologna, 2014.
- VERSIGLIONI M., *Prova e studi di settore*, Milano, 2007.
- VIOTTO A., *I poteri di indagine dell'amministrazione finanziaria*, Milano, 2002.
- WARD J. S. – BARKER A., *Undefined by data: a survey of a Big data definitions*, University of St. Andrews, UK, 20 september 2013.
- WEINBERGER D., *Everything is mischellaneous: the power of the new digital disorder*, Henry Holt and Company, New York, 2008, tr. it., *Elogio del disordine. Le regole del nuovo mondo digitale*, Rizzoli, Milano, 2010.
- WEINBERGER S., *Terrorist «Pre-Crime» Detector Field Tested in United States*, in *Nature*, 27 maggio 2011.
- ZACCARIA R., *Fuga dalla legge? Seminari sulla qualità della legislazione*, Brescia, 2011.