

## Un modello decentralizzato a garanzia dei diritti fondamentali dell'individuo nei processi decisionali data-driven

Corrado Ameli<sup>[1]</sup>, Simone Mallardo<sup>[2]</sup>

*I paragrafi I, II, V si devono a Simone Mallardo; i paragrafi III, IV a Corrado Ameli.*

(ABSTRACT)

*Con l'avvento della società dell'informazione e della digitalizzazione, i dati hanno iniziato ad assumere un ruolo di primaria importanza nell'ambito della decisione politica e/o amministrativa. Oggi, l'analisi di grandi quantità di dati permette di predire, seppur con margini di errore, l'avverarsi di scenari che, a loro volta, possono essere analizzati al fine di attuare piani probabilisticamente più efficaci rispetto agli obiettivi stabiliti. Tuttavia, modalità scorrette e poco trasparenti di acquisizione, filtraggio ed elaborazione dei dati possono produrre bias nell'attuazione delle strategie politiche, che rischiano di divenire fallaci. Decisioni politiche oggettivamente opinabili possono così diventare "giustificabili" sulla base di dati che non rispecchiano la realtà. In questo contributo viene proposto un modello decentralizzato che regola l'intero iter di formazione dei dati in modo aperto e trasparente permettendo così di sindacare la logicità e la ragionevolezza delle decisioni assunte proprio sulla base di quei dati. Il modello proposto ha come fine ultimo quello di permettere al cittadino l'accesso ad informazioni che siano il più genuine possibili, tutelandolo dalla disinformazione e dalla formazione di opinioni politiche false, distorte o comunque non aderenti alla realtà.*

### I. Introduzione.

Il rapporto tra *data science* e diritto<sup>[3]</sup> è sempre più stretto: la prima occupa ormai spazi sempre più significativi della nostra quotidianità, incentivando una domanda di regolazione da parte del secondo. In questa prospettiva, particolare rilevanza hanno assunto i processi decisionali basati sui dati<sup>[4]</sup>; si pensi, ad esempio, al ricorso ai processi decisionali automatizzati (*data-driven decision making*) per fronteggiare il contenimento dei contagi nell'ambito della pandemia Covid-19<sup>[5]</sup>.

In questa prospettiva, lo scritto raccoglie la sfida lanciata da Zeno-Zencovich concernente la necessità di introdurre presso gli ordinamenti giuridici processi di formazione e analisi dei dati che siano verificabili dal cittadino, al fine di evitare impostazioni arbitrarie da parte di poteri pubblici e privati<sup>[6]</sup>.

Questo tema verrà approfondito concentrandosi, in particolare, sulle problematiche concernenti l'iter di formazione dei dati, suggerendo, altresì, uno specifico modello alternativo di carattere decentralizzato. In quest'ottica - adottando una impostazione che si basa sull'integrazione tra riflessione giuridica e *data analysis* - si indagheranno le impostazioni a livello europeo e nazionale che, avallando un approccio *open government*, depongono a favore, o comunque non si ritiene ostacolino, l'introduzione di un paradigma decentralizzato nell'iter di formazione dei dati (*data life cycle*); successivamente si analizzeranno gli obiettivi, i vantaggi e gli svantaggi sottesi all'adozione del modello; ed infine, si proporrà l'introduzione del modello di formazione, filtraggio, elaborazione e analisi dei dati caratterizzato da un meccanismo di reciproco controllo fra i soggetti coinvolti, a beneficio diretto della qualità<sup>[7]</sup> dei dati nel rispetto del principio di trasparenza<sup>[8]</sup> e democraticità sanciti dal nostro ordinamento e, da ultimo, nel pieno rispetto del diritto alla buona amministrazione di cui agli artt. 97 Cost. e 41 della Carta di Nizza.

## **I.1. - L'attuazione politica e il quadro giuridico dell'UE in riferimento alla circolazione e alla qualità dei dati.**

Le trasformazioni radicali che hanno coinvolto l'economia, la cultura, il lavoro, il commercio, i settori produttivi di beni e servizi nell'ultimo ventennio sono da ricondurre, tra l'altro, al progresso tecnologico, vero e proprio motore del processo di modernizzazione<sup>[9]</sup> della società il quale, però, necessita di essere governato per poter perseguire l'interesse pubblico e lo sviluppo del benessere della collettività<sup>[10]</sup>.

In questa prospettiva, una delle priorità delle istituzioni europee nel breve periodo<sup>[11]</sup> è favorire un coerente sviluppo economico basato sui dati all'interno degli Stati membri. Come si vedrà, questo obiettivo può essere raggiunto adottando diversi modelli di formazione dei dati; ciò nondimeno, la necessità di favorire un'impostazione che si qualifichi per la loro gestione *open* ha ormai assunto una rilevanza cruciale. Per chiarire questo punto, si analizzerà la principale normativa europea sul tema; questo passaggio ci consentirà, successivamente, di identificare le differenze con l'approccio seguito nel nostro paese.

Il primo atto normativo dell'UE in materia di dati<sup>[12]</sup> è la [Direttiva 2003/98/CE](#) relativa al riutilizzo dell'informazione del settore pubblico che, all'art. 3, ha stabilito il principio fondamentale per il quale "*gli Stati membri provvedono affinché, ove sia permesso il riutilizzo di documenti in possesso degli enti pubblici, questi documenti siano riutilizzabili [...]*". La Direttiva è volta ad armonizzare la disciplina sul riutilizzo dei dati e delle informazioni<sup>[13]</sup> del settore pubblico; in quest'ottica, il "riutilizzo" dei dati si riferisce sia all'aspetto soggettivo che a quello teleologico: potranno riutilizzare i dati enti privati, altre pubbliche amministrazioni, società, persone giuridiche o fisiche ed anche macchine, intelligenze artificiali, e qualunque altra applicazione che necessiti di dati per funzionare, per qualunque scopo - sia esso commerciale o meramente conoscitivo - purché nel rispetto della normativa in materia<sup>[14]</sup>.

A questo primo provvedimento si deve aggiungere la comunicazione [COM \(2011\) 882](#) "Dati aperti, un motore per l'innovazione, la crescita e una *governance* trasparente" mediante la quale si è sancito che l'efficienza delle amministrazioni e la partecipazione dei cittadini alla vita sociale e politica possono essere migliorati attraverso una più ampia circolazione dei dati proponendo, del pari, tre interventi:

- uniformare il quadro normativo sul riutilizzo dei dati mediante un'opera di revisione

- della direttiva del 2003 e l'adozione di altri atti legislativi;
- sostenere e finanziare<sup>[15]</sup> una strategia *open data* tramite progetti di ricerca, creando anche portali per l'accesso ai dati europei<sup>[16]</sup>;
  - incrementare il coordinamento e la condivisione dei dati fra gli Stati membri.

La comunicazione del 2011 è una forte “presa d’atto” sulla necessità di implementare una strategia decentralizzata nella formazione e gestione dei dati; secondo le istituzioni europee, infatti, questo approccio permetterebbe di scongiurare “tendenze monopolistiche che si fondano su dati provenienti da un’unica fonte”; in questa prospettiva si suggerisce di favorire l’applicazione del principio della riutilizzabilità delle informazioni detenute dal settore pubblico per qualunque lecito fine, di ridurre i costi addebitabili per la circolazione delle informazioni e di ampliare il campo di applicazione della direttiva a biblioteche, archivi e musei.

Queste proposte sono state effettivamente adottate nella [direttiva 2013/37/UE](#) sulle politiche relative all’apertura, alla circolazione e al riutilizzo dei dati detenuti dalle pubbliche amministrazioni per favorire la crescita della conoscenza<sup>[17]</sup> e dell’economica. La direttiva 37/2013 è di cruciale importanza in quanto afferma che il riutilizzo dei dati, delle informazioni e dei documenti, promuovendo la trasparenza e la responsabilizzazione, grazie al feedback fornito dal riutilizzatore del dato, permette anche all’originario detentore di migliorare la loro qualità<sup>[18]</sup>.

Nello stesso solco, il 2 luglio 2014, la commissione europea ha approvato la comunicazione [COM \(2014\) 442](#) “Verso una florida economia basata sui dati”, accogliendo l’invito del Consiglio europeo dell’anno precedente di articolare proposte operative dirette alla creazione di un mercato unico dei *big data*<sup>[19]</sup>. La comunicazione considera i dati “*il fulcro dell’economia e della società della conoscenza del futuro*” specificando che, attraverso la loro analisi, è possibile conseguire un miglioramento dei processi decisionali contribuendo, tra l’altro, a “*generare nuove idee o soluzioni o a prevedere gli eventi futuri con maggiore precisione*”<sup>[20]</sup>. D’altra parte, questo atto è particolarmente significativo poiché rivela il tentativo di creare un mercato unico digitale fondato sulla *data economy*; un obiettivo, questo, che necessita però di *dataset* di buona qualità, ossia affidabili ed interoperabili<sup>[21]</sup>.

In occasione del Consiglio europeo del 26 e 27 giugno 2014 sono state discusse le priorità chiave<sup>[22]</sup> che l’agenda strategica per l’UE avrebbe dovuto contenere rispetto al periodo 2015-2020. Così, presso l’allegato 1 alle conclusioni del Consiglio europeo, denominato “Agenda strategica per l’Unione in una fase di cambiamento”<sup>[23]</sup>, si fa esplicito riferimento ad una “corsa all’innovazione nell’era digitale” che deve incentivare gli stati e le organizzazioni internazionali ad anticipare le occasioni che emergeranno dal progresso tecnologico, ad esempio, mediante il perfezionamento del mercato unico digitale entro l’anno successivo.

Le proposte del Consiglio europeo sono state accolte dalla Commissione EU che nel 2015 ha presentato la comunicazione [COM \(2015\)192](#) “Strategia per il mercato unico digitale in Europa”<sup>[24]</sup>. La comunicazione stabilisce che è necessario superare la rigidità delle regolamentazioni nazionali in favore di un approccio più dinamico e aperto il quale, a sua volta, poggia sulla massimizzazione del “potenziale di crescita dell’economia digitale europea” attraverso investimenti proprio nei *big data* considerati dei veri e propri “volano”

della crescita economica e della competitività europea.

Ciò posto, al netto delle affermazioni delle istituzioni europee, sono molti gli ostacoli che frenano lo sviluppo di un mercato unico digitale dei dati e l'elaborazione di nuove tecnologie e servizi digitali. Tali impedimenti possono essere ricondotti alla frammentaria disciplina normativa, agli ostacoli di natura tecnica, alla poca chiarezza delle regole giuridiche sull'uso dei dati, ma - soprattutto - alla "mancanza di sistemi e servizi aperti e interoperabili e di portabilità dei dati fra servizi"<sup>[25]</sup>. Si comprendono, così, le ragioni della strategia europea: la creazione di uno spazio unico in cui i dati non personali possano circolare liberamente. In tal ottica, la commissione, all'interno del punto 5 della comunicazione, indica la necessità di impegnarsi per migliorare la qualità dei dati. Così, la qualità dei dati e l'impiego di un modello aperto si pongono come presupposti di un'unica strategia che - d'altra parte - implica una maggiore attenzione delle istituzioni europee alle fasi antecedenti alla circolazione dei dati<sup>[26]</sup>.

Nel 2018, anche grazie al [Regolamento \(UE\) 2016/679](#) e alla [COM \(2017\) 9](#) "Costruire un'economia dei dati europea"<sup>[27]</sup>, le politiche euro-unitarie hanno fatto un ulteriore passo in avanti mediante l'approvazione della comunicazione [COM \(2018\) 237](#) denominata "L'intelligenza artificiale per l'Europa"<sup>[28]</sup>, alla quale farà seguito, qualche mese dopo, il [Regolamento \(UE\) 2018/1807](#) dedicato alla libera circolazione dei dati non personali nell'UE<sup>[29]</sup> mediante il quale si è stabilito che "gli obblighi di localizzazione di dati sono vietati a meno che siano giustificati da motivi di sicurezza pubblica nel rispetto del principio di proporzionalità" vincolando gli Stati membri ad abrogare qualsiasi obbligo di localizzazione dei dati entro il 30 maggio 2021<sup>[30]</sup>.

Completa il quadro la [Direttiva 2019/1024/UE](#) che sancisce il principio del riutilizzo dei dati e delle informazioni detenute da enti pubblici in favore di qualsiasi persona fisica o giuridica<sup>[31]</sup> secondo una logica G2B<sup>[32]</sup> (*government-to-business*). Tale direttiva garantisce che i detentori pubblici di dati li mettano a disposizione in quantità sempre maggiore per l'utilizzo non solo da parte delle imprese ma anche della società civile e della comunità scientifica<sup>[33]</sup>.

Da ultimo, questi interventi normativi sono culminati nel 2020 con l'approvazione delle seguenti comunicazioni:

- Libro Bianco sull'intelligenza artificiale ([COM\(2020\) 65](#));
- Una strategia europea per i dati ([COM\(2020\) 66](#));
- Plasmare il futuro digitale dell'Europa ([COM\(2020\) 67](#)).

Tra questi provvedimenti, assume particolare interesse il secondo che contiene un'ottima sintesi delle misure adottate fino ad oggi all'interno dell'UE, e descrive puntualmente quali saranno le strategie per i prossimi anni. Ribadendo che i dati sono i nuovi protagonisti dell'innovazione digitale, la Commissione afferma la necessità che la metodologia di raccolta e utilizzo degli stessi non può prescindere dal rispetto dei diritti fondamentali dei cittadini europei, fra cui la possibilità di adottare decisioni migliori sulla base delle informazioni ottenute dai dati di alta qualità<sup>[34]</sup>. In questa ottica, la comunicazione specifica l'importanza del ruolo del cittadino nella società dell'informazione considerandolo protagonista della gestione e dell'utilizzo dei dati attraverso l'adozione di soluzioni completamente decentrate<sup>[35]</sup>. Si può affermare, dunque, che l'obiettivo ultimo dell'UE è la creazione di una

società ove sia possibile prendere decisioni consapevoli e sviluppare servizi migliori proprio grazie ai dati (*data driven innovation*)<sup>[36]</sup> che siano aperti, condivisibili e consultabili da chiunque rispettando il principio di trasparenza<sup>[37]</sup>.

Dall'analisi del quadro normativo europeo si ricava l'apprezzabile sforzo volto allo sviluppo della civiltà della conoscenza e della *data economy* attuando uno spostamento del *focus* dalla messa a disposizione dei dati alla loro qualità (caratteristica inscindibilmente collegata al *data life cycle*) perché risulterebbe inutile avere a disposizione una enorme quantità di dati pregiudicati da *bias*: a quantità deve corrispondere qualità<sup>[38]</sup>.

## **I.II. Il contesto italiano: il legislatore contempla l'iter di formazione del dato?**

Anche se all'interno dell'ordinamento giuridico italiano non è possibile rinvenire normativa che regoli esplicitamente i *big data* e il loro procedimento di formazione, vi sono alcune disposizioni che forniscono spunti di riflessione utili ai nostri fini.

Un primo riferimento utile dal quale partire in tema di dati è il codice dell'amministrazione digitale (CAD), istituito con d.lgs. n. 82/2005 e successive modificazioni<sup>[39]</sup> (d.l. n. 179/2012, d.lgs. n. 179/2016, d. lgs n. 217/2017 e d.l. n. 76/2020).

In particolare, l'art. 50 CAD, ponendosi in linea di continuità con l'approccio di Bruxelles, stabilisce che *"i dati delle pubbliche amministrazioni sono formati, raccolti, conservati, resi disponibili e accessibili con l'uso delle tecnologie dell'informazione e della comunicazione (TIC) che ne consentano la fruizione e riutilizzazione, alle condizioni fissate dall'ordinamento, da parte delle altre pubbliche amministrazioni e dai privati [...]"*<sup>[40]</sup>. Si noti che la disposizione cita espressamente l'iter di formazione ancor prima della fase di raccolta dati.

Un riferimento alla formazione dei dati è altresì contenuto nel d.l. n. 179/2016 che con l'art. 1 c. 1 lett. g) ha inserito, all'art. 1 del CAD lett. cc), la definizione di "titolare del dato" individuandolo in colui "che ha originariamente formato per uso proprio o commissionato ad altro soggetto il documento che rappresenta il dato, o che ne ha la disponibilità"<sup>[41]</sup>. In tal modo, ai sensi del CAD, il documento è un atto che contiene la rappresentazione di dati giuridicamente rilevanti<sup>[42]</sup>. Ora, benché la disposizione riferisca l'attività di formazione al documento che rappresenta il dato e non al dato stesso, essa qualifica colui che redige il documento come il titolare del solo dato e non del documento; di conseguenza si ritiene che il processo di formazione possa riferirsi anche al dato oltre che al documento<sup>[43]</sup>.

A queste disposizioni devono poi aggiungersi altre previsioni normative concernenti la formazione dei dati. La prima è l'art. 41 c. 2 CAD: *"la pubblica amministrazione titolare del procedimento raccoglie in un fascicolo informatico gli atti, i documenti e i dati del procedimento medesimo da chiunque formati"*. Tale disposizione è un ulteriore elemento che depone a favore del fatto che, normativamente, un processo di formazione dei dati è contemplato. Lo stesso ragionamento vale per l'art. 49 CAD che, stabilendo il divieto per gli addetti alla trasmissione di atti, dati e documenti di prenderne conoscenza, contempla la formazione di dati con strumenti informatici. Infine, vi è l'art. 52 c. 3 CAD che fa riferimento a prodotti e servizi che possono formare, raccogliere e gestire dati.

L'analisi della normativa nazionale consente di specificare qualche contrappunto di sintesi.

In primo luogo, emerge una differenza lessicale che, al contempo, provoca un disallineamento anche concettuale. In particolare, il CAD fa un mero riferimento, all'interno di alcune sue disposizioni, ad una fase di formazione dei dati della quale ne riconosce l'esistenza ma non la regola; ancor più laconica è la normativa europea che non fa alcun riferimento all'iter di formazione dei dati se non indirettamente quando tratta della loro qualità. Nonostante ciò, la *data quality* è caratteristica sempre più imprescindibile e per tale motivo è necessario che il procedimento di formazione sia corretto dal punto di vista tecnico e non condizionato da interessi di carattere personale, economici o politici<sup>[44]</sup>.

Assodata la laconicità della regolamentazione, nasce l'esigenza di integrare i riferimenti normativi con alcune determinazioni contenute nell'indagine conoscitiva delle autorità garanti sui *big data*.

L'autorità per le garanzie nelle comunicazioni (AGCOM), unitamente all'autorità garante della concorrenza e del mercato (AGCM) e all'autorità garante per la protezione dei dati personali ha avviato, il 30 maggio 2017, con la [Delibera n. 217/17/CONS](#) una indagine conoscitiva sui *big data*, conclusasi il 27 novembre 2019<sup>[45]</sup>.

L'indagine si è soffermata, in particolare, sulla c.d. "filiera dei *big data*" (*big data life cycle*) che è il processo di creazione ed estrazione della conoscenza dai dati; in esso "è possibile enucleare, sul piano logico (con possibili ricadute anche su quello giuridico) tre ordini principali di attività: i) la raccolta, che a sua volta si articola in generazione, acquisizione e memorizzazione, ii) l'elaborazione, che coinvolge attività di estrazione, integrazione e analisi, iii) l'interpretazione e l'utilizzo"<sup>[46]</sup>. I dati creati da questo procedimento di formazione possono essere interpretati ed utilizzati nei processi decisionali<sup>[47]</sup>. Ciò premesso, la tipologia di soggetti che ruotano attorno ai dati e che si occupano dei relativi processi di formazione e la loro posizione giuridica è un elemento centrale. Sul punto, l'indagine conoscitiva rileva come la titolarità dei dati appartenga a pochi "player internazionali di grandi dimensioni": vi è una forte centralizzazione dei dati in poche multinazionali private oppure nelle pubbliche amministrazioni che li detengono. Questa situazione di fatto ha portato le istituzioni comunitarie e le autorità garanti ad interessarsi delle politiche attinenti alla titolarità e alla formazione dei dati per evitare che un approccio troppo centralizzato possa determinare un accumulo eccessivo di informazioni nelle mani di pochi.

In quest'ottica è stata proposta l'implementazione di modelli alternativi. Le ragioni che spingono le autorità a proporre un modello diverso da quello attuale non sono esclusivamente riconducibili al principio di trasparenza<sup>[48]</sup> e di democraticità delle decisioni, ma anche ai profili concernenti la qualità dei dati grazie all'adozione di procedure e linee-guida nonché l'affidabilità dei criteri di analisi dei dati selezionati<sup>[49]</sup>. In questo senso, la garanzia del principio di qualità dei dati è fondamentale onde evitare ripercussioni negative sulla correttezza dell'azione amministrativa: come si evince dalla relazione, infatti, "il trattamento di informazioni di scarsa qualità (ad es. incomplete, inesatte o risalenti) come pure l'impiego di tecniche di ricerca (algoritmi) che presentano errori di configurazione, sono suscettibili di condurre ad inferenze scorrette che si ripercuotono, in ultima analisi, sul processo (conoscitivo o decisionale) posto in essere e quindi su singoli o gruppi (più o meno ampi) di individui".

Per superare queste criticità l'autorità garante propone la previa sperimentazione<sup>[50]</sup> degli

esiti delle analisi dei *big data* su scala ridotta e, in secondo momento, applicare questa metodologia ad un contesto più ampio.

## **II. Obiettivo: l'attuazione di un approccio *open government*.**

Pur riconoscendo l'utilità e la necessità dell'approccio sperimentale, come proposto dalle autorità garanti, il quale consentirebbe una verifica sui risultati della sperimentazione e sulle sue fasi, emerge la possibilità di seguire un'altra impostazione volta a garantire la qualità dei dati: l'adozione di un modello decentralizzato nell'iter di formazione degli stessi.

In generale, infatti, come rilevato da autorevole dottrina<sup>[51]</sup>, gli approcci caratterizzati da una forte centralizzazione difficilmente si coniugano con esigenze di trasparenza e di proporzionalità delle decisioni: la concentrazione di potere - e di detenzione dei dati - può creare, in assenza di specifiche garanzie volte a rendere verificabili i principali passaggi concernenti la formazione dei dati, le condizioni per un cattivo esercizio del potere<sup>[52]</sup>.

Del resto, rispetto a questo approccio, le stesse autorità garanti hanno suggerito l'opportunità di adottare "architetture alternative" anche a causa dell'emersione di possibili asimmetrie tra i detentori dei dati e gli utenti che li mettono originariamente a disposizione<sup>[53]</sup>.

Anche alla luce di queste considerazioni, le autorità garanti propongono di prendere le distanze da modelli caratterizzati da forte centralizzazione per approdare ad un modello decentralizzato ove ciascun utente sia in grado di controllare - se non addirittura partecipare - alla gestione dei propri dati decidendone la sorte<sup>[54]</sup>.

Gli stessi benefici - in termini di democraticità e trasparenza - di un approccio decentralizzato<sup>[55]</sup> possono replicarsi anche nella loro fase di formazione che non sarebbe più concentrata nelle mani di un unico soggetto (pubblico o privato che sia) bensì verrebbe separata fra più operatori che si pongono in rapporto di reciproco controllo.

In termini generali, l'adozione di un metodo decentralizzato permette l'implementazione di un modello di amministrazione ispirato altresì all'*open government*<sup>[56]</sup>, vale a dire *"un'impostazione che predica l'apertura della pubblica amministrazione ai cittadini sia sotto il profilo della trasparenza sia per quel che attiene alla partecipazione al processo decisionale anche grazie all'utilizzo delle nuove tecnologie della comunicazione"*<sup>[57]</sup>; tale modello si fonda su tre pilastri: trasparenza (dati disponibili e condivisibili), partecipazione (coinvolgimento dei cittadini e dei privati nel sistema decisionale), collaborazione (cooperazione tra diversi livelli di governo e gli attori privati)<sup>[58]</sup>.

Coinvolgendo le imprese private che collaborano con la pubblica amministrazione si realizza il principio di trasparenza<sup>[59]</sup> e partecipazione comportando una fisiologica e proficua "contaminazione" tra mondo pubblico e privato<sup>[60]</sup> nell'iter di formazione dei dati.

Il modello decentralizzato si affianca agli strumenti tradizionali previsti dal diritto amministrativo per rendere effettivo il principio di trasparenza quali l'accesso, la partecipazione e l'obbligo di motivazione<sup>[61]</sup>. Inoltre, attraverso l'attuazione di questo modello, le pubbliche amministrazioni incentivano il dialogo con gli enti privati, rendendoli protagonisti dell'iter di formazione dei dati e, indirettamente, delle decisioni che

l'amministrazione adotterà sulla base di quei dati. Si realizza così il principio di partecipazione<sup>[62]</sup> dei privati, seppur indirettamente, alle decisioni pubbliche.

D'altra parte, non si può prescindere dal fatto che utilizzando i *big data* le logiche di quantità e di correlazione rischiano di sopraffare le caratteristiche attinenti alla loro qualità: necessarie per garantire certezza, affidabilità e fiducia<sup>[63]</sup>.

Esaurita la trattazione riguardante il quadro normativo e programmatico si procede con la seconda parte dell'elaborato che presenta il nuovo modello decentralizzato nell'iter di formazione dei dati (*data life cycle*) illustrandone il funzionamento, la concreta applicazione e i relativi vantaggi e svantaggi.

### III. Metodologia.

#### III.I. L'oggetto del modello: il dato informatico.

Al fine di stabilire un vocabolario univoco, riportiamo definizioni di concetti base inerenti al dato informatico.

Il dato è una rappresentazione originaria, non interpretata, di un evento o di un fenomeno effettuata attraverso dei simboli<sup>[64]</sup>. L'informazione viene ricavata dal dato tramite un processo di interpretazione. L'intero processo che porta alla creazione dei dati e alla relativa interpretazione viene chiamato iter del dato (o filiera del dato<sup>[65]</sup>).

Dividiamo l'iter del dato in sei partizioni:

1. *Pianificazione*: in questa fase viene redatto un documento in cui si definiscono le motivazioni e gli obiettivi dello studio. Vengono decise le modalità di acquisizione del dato, specificando eventuali protocolli che devono essere rispettati nelle fasi successive al fine di minimizzare il rischio di *bias*<sup>[66]</sup>.
2. *Acquisizione*: i dati vengono acquisiti secondo le modalità descritte nella fase di pianificazione.
3. *Ispezione*: viene effettuato un controllo che certifica la qualità del dato raw<sup>[67]</sup>. Nel caso in cui alcune misurazioni o osservazioni risultino incompatibili con gli standard prefissati dalla fase di pianificazione, o palesano evidenti difformità dalla realtà, essi vengono rimossi.
4. *Elaborazione e Filtraggio*: i dati *raw* vengono elaborati in un formato più fruibile (*processed data*). Questa fase può prevedere conversioni di unità di misura, estrapolazione di ulteriori dati dai dati già acquisiti o filtraggio di *noise*<sup>[68]</sup>.
5. *Integrazione*: nel caso in cui la fase di acquisizione abbia dato origine a dati di diverso profilo<sup>[69]</sup>, in questa fase essi vengono unificati (o integrati) di modo da ottenere una collezione di dati pronta all'analisi.
6. *Analisi*: vengono utilizzati metodi di inferenza statistica per calcolare parametri, differenze o previsioni con i relativi margini di errore.

Le fasi 3, 4 e 5 possono essere ripetute, ignorate e/o scambiate di ordine a seconda del tipo di scenario.

### III.II. Modello Decentralizzato Anonimizzato.

Del pari, prima di descrivere il Modello Decentralizzato Anonimo (MDA) è necessario fornire alcune definizioni di vocaboli utilizzati nelle successive sezioni:

- *Ente Richiedente*: l'ente pubblico che richiede l'indagine<sup>[70]</sup>.
- *Ente Privato*: ente privato certificato che ha la possibilità di offrire la sua disponibilità per partecipare ad un'indagine offerta<sup>[71]</sup>.
- *Ente Operativo*: uno dei sei enti privati certificati che viene eletto per partecipare allo sviluppo di un'indagine. Ad ogni ente operativo spetta una fase dell'iter di formazione dei dati.
- *Ente Esecutivo*: l'ente operativo incaricato di portare a termine una precisa fase dell'iter.
- *Ente Controllore*: l'ente operativo che non riveste il ruolo di ente esecutivo durante una particolare fase, il quale si occupa di controllare e approvare (o respingere) l'operato dell'ente esecutivo<sup>[72]</sup>.
- *Piattaforma*: strumento telematico che interfaccia l'ente richiedente agli enti privati. Il codice sorgente che implementa il funzionamento della piattaforma è open source. Nessuno degli enti ha modo di modificare scelte prese dalla piattaforma né alterarne il funzionamento.
- *Credenziale Anonima*: meccanismo<sup>[73],[74]</sup> per il quale, ad ogni indagine, un ente operativo rimane anonimo nei confronti degli altri enti operativi e dell'ente richiedente.

Ciò posto, il MDA è articolato in quattro fasi:

- Fase di selezione;
- Fase di discussione;
- Fase di contrattazione;
- Fase di esecuzione;

#### **Fase di Selezione**

L'ente richiedente inserisce nella piattaforma l'indagine con l'impegno di redigere il documento di *fundamentum* (ref. III.III). Gli enti privati vengono notificati dell'inserimento di un'indagine nella piattaforma. Ogni ente privato può decidere, data la sua disponibilità e la sua esperienza legata alla tipologia di indagine richiesta, se partecipare o meno. Nel momento in cui un ente privato aderisce, la piattaforma gli fornisce una credenziale anonima che ha lo scopo di non rivelare l'identità dell'ente. Il numero di enti privati che hanno aderito all'indagine rimane oscurato per tutte le parti. Terminato il tempo di selezione dei partecipanti, la piattaforma estrae casualmente sei enti privati, uno per ogni fase dell'iter del dato (ref. III.I). In questa fase, gli enti operativi non sono ancora a conoscenza dell'assegnazione delle fasi.

#### **Fase di Discussione**

Gli enti operativi possono interagire con l'ente richiedente attraverso la piattaforma. Ogni comunicazione al di fuori della piattaforma è vietata. La piattaforma tiene traccia di ogni messaggio scambiato tra i soggetti. Non sono contemplate né ammesse modalità di

interazione privata, né tra enti operativi, né tra un singolo ente operativo e l'ente richiedente; questa regola è estesa anche alle fasi successive. In questa fase vengono argomentati tutti i punti stilati nel documento di *fundamentum* attraverso una interazione aperta tra le parti. La fase viene conclusa quando tutte le parti si ritengono soddisfatte.

### **Fase di Contrattazione**

La piattaforma assegna ad ogni ente operativo un ruolo, che corrisponde ad una delle fasi dell'iter del dato. Ogni ente operativo inserisce sulla piattaforma le proprie richieste economiche, dividendo la somma tra l'ammontare da ricevere a inizio indagine e l'ammontare atteso a fine indagine. L'ente richiedente può visualizzare sulla piattaforma solo la somma dei costi di tutti gli enti operativi (divisi per ammontare a inizio e fine indagine), non la ripartizione per singolo ente. Se l'ente richiedente è soddisfatto dell'offerta, si passa alla fase di esecuzione. Se l'ente richiedente ritiene che l'offerta non sia adeguata, esso avvisa gli enti operativi, i quali avranno la possibilità di mantenere la loro proposta inalterata, ripartire diversamente l'ammontare a inizio e a fine indagine o di abbassare il valore dell'offerta. Non è possibile per gli enti operativi aumentare il valore totale dell'offerta, né incrementare il valore della somma spettante a inizio indagine. La contrattazione può ripetersi fino ad un massimo di 3 volte. Se dopo la terza contrattazione l'ente richiedente non approva il costo dell'indagine, la procedura ricomincia dalla fase di selezione. In tal caso, tutte le comunicazioni effettuate durante la fase di discussione vengono rese disponibili nella prossima iterazione.

### **Fase di Esecuzione**

Partendo dalla fase di pianificazione dell'iter del dato, gli enti operativi lavorano attivamente allo sviluppo dell'indagine, in veste di esecutore o di controllore al cambiare di ogni fase.

Ogni ente esecutivo ha l'obbligo di effettuare il lavoro a regola d'arte, utilizzando metodi scientifici ben consolidati in letteratura. Ogni esecuzione di una fase deve essere documentata propriamente seguendo le indicazioni riportate nel Documento di *Relatio* (ref. III.IV).

Ogni ente controllore ha l'obbligo di revisionare il lavoro di ogni ente esecutivo e di approvare o respingere, con motivazioni di base scientifica, l'operato degli ultimi. Una fase viene considerata conclusa quando approvata da tutti gli enti controllori. Nel caso ci sia almeno un ente controllore che respinge l'esecuzione di una fase, l'ente esecutivo ha il compito di rielaborare l'operato. Nel caso in cui l'ente esecutivo sia fermamente convinto della bontà dell'operato, esso può sottomettere di nuovo a revisione lo stesso elaborato. Il numero di rielaborazioni massime di una fase viene stabilito dall'ente richiedente nel Documento di *Fundamentum*. Nel caso in cui solo la minoranza degli enti controllori (massimo due su cinque) continui a respingere l'operato di un ente esecutivo, la decisione finale di approvazione/rigetto dell'operato viene affidata ad un altro ente privato che non ha avuto alcun tipo di relazione con l'indagine in corso ma che ha precedentemente aderito alla fase di selezione. Nel caso in cui la maggioranza degli enti controllori (almeno tre su cinque) continua a respingere l'operato di una fase, quest'ultima viene riaffidata ad un altro ente privato (segue lo stesso criterio del punto precedente).

Una fase dell'iter del dato può essere soggetta a integrazione di ulteriore lavoro durante lo sviluppo di fasi successive; permane comunque l'obbligo di stilare ad ogni integrazione il Documento di *Relatio*.

### **III.III. Documento di *Fundamentum*.**

Il Documento di *Fundamentum* è redatto dall'ente richiedente e ha come scopo quello di fornire tutte le informazioni necessarie affinché gli enti operativi possano essere in grado di comprendere appieno la tipologia di lavoro richiesta.

Il documento deve comprendere:

- Le motivazioni che giustificano la necessità di effettuare una nuova indagine.
- Gli obiettivi che si vogliono perseguire con l'ausilio dell'indagine, stilati in maniera diretta e specifica.
- Una lista che delinea la priorità di ogni obiettivo.
- Il livello di confidenza atteso per ogni obiettivo.
- Il livello di scala dell'analisi.
- Limiti temporali e numero di rielaborazioni massime.

Il documento può eventualmente comprendere, nel caso si applichi:

- Riferimenti a indagini passate per le quali urge una necessità di integrazione, ampliamento o confronto.
- Istruzioni sulla pseudoanonimizzazione e l'immagazzinamento di possibili dati sensibili.
- Istruzioni e referenze per collaborazioni con enti pubblici.
- Istruzioni e referenze per collaborazioni con enti esteri.
- Accesso a basi di dati private già disponibili, con relativa documentazione.

Il documento non deve fare riferimento a scelte di carattere tecnico inerenti all'indagine quali (ed in modo non esaustivo): approcci qualitativi o quantitativi, studi controllati o randomizzati, esperimenti aperti o in cieco, metodi di campionamento della popolazione, protocolli di categorizzazione del dato, riferimenti a metodi statistico-inferenziali specifici.

### **III.IV. Documento di *Relatio*.**

Il Documento di *Relatio* viene redatto dall'ente operativo alla fine di una fase e accompagna lo scambio di dati nella piattaforma. Il Documento di *Relatio* si presenta come una lista di linee guida, suddivise per fasi dell'iter, e ha lo scopo ultimo di documentare l'indagine e di delineare un profilo qualitativo del lavoro svolto. Il Documento è stato stilato seguendo le linee guida riportate nei seguenti studi<sup>[75]-[76]</sup>, aggiungendo e rielaborando materiale per i fini del manoscritto.

Legenda simboli:

- Richiede riferimento a un dato immagazzinato. □
- Richiede riferimento a codice sorgente / regole di implementazione. □
- Richiede riferimento a documento precedente disponibile nella piattaforma. □

Per ogni punto, l'ente operativo ha il compito di decidere se il punto in questione è applicabile ed è stato applicato (con riferimenti stilati nella leggenda simboli), o non è applicabile. Le linee guida del Documento di Relatio sono:

- Acquisizione

1. Origine: il dato originale (o raw), deve essere immagazzinato e descritto prima di qualsiasi trasformazione, modifica. □
2. Metodo di acquisizione: viene descritta la tecnologia utilizzata per registrare valori in formato elettronico. □
3. Protocollo di acquisizione: si descrive come il dato è stato ottenuto. □□
4. Scopo di utilizzo: descrizione del contesto per il quale i dati sono stati originati. □□
5. Origine esterna ed enti coinvolti: descrizione degli enti che hanno collaborato nel fornire un dato già acquisito e immagazzinato precedentemente. □□

- Ispezione

1. Validazione del dato: descrizione di come l'immagazzinamento è stato effettuato. □□
2. Validazione dell'acquisizione: descrizione di come sono stati effettuati dei controlli per verificare la bontà della fase di acquisizione. □□
3. Validazione di consistenza: descrive come sono stati effettuati i controlli atti a confermare la consistenza dei dati, ad esempio, ed in maniera non esaustiva: consistenza di cardinalità', co-occorrenza, co-misurazione, precision loss. □□

- Integrazione

1. Struttura del database: Una descrizione di come diversi dati sono strutturati e collegati tra loro. Include informazioni sulla natura del dato (tipo, intervallo o valori nel caso di variabili categoriche). Include regole che stabiliscono se un'assegnazione di valori è opzionale o meno. □
2. Dizionario del dataset: Una descrizione della semantica del dato. □

- Processing

1. Conversione di valori: descrive come dei dati sono stati trasformati/convertiti in un altro formato. □□
2. Creazione di variabili costruite: descrive come dei dati sono stati processati in modo da ottenere una nuova variabile. □□

- Analisi

1. Statistica descrittiva: descrive come per ogni variabile sono stati calcolati valori di statistica descrittiva (e.g. media, frequenza) e test di *goodness of fit* atti a inferire proprietà di distribuzioni del dato. □□
2. Statistica inferenziale: descrive come sono stati calcolati valori o stime da processi di statistica inferenziale. □□

#### **IV. Discussione.**

La discussione della metodologia è divisa in due parti, nella prima parte vengono sviluppate e giustificate le scelte adottate nella costruzione della metodologia, mentre nella seconda

parte vengono analizzati gli svantaggi derivanti dall'attuazione del MDA.

#### **IV.I. Commenti sulla metodologia.**

L'intera metodologia è ispirata al concetto del *Distributed Ledgers*. Non vi è una amministrazione centrale; ogni singolo nodo può operare anonimamente e ha il compito di controllare l'operato degli altri. Studi hanno dimostrato i benefici derivanti dall'utilizzo di approcci decentralizzati<sup>[77],[78]</sup>, senza però fornire un'implementazione utilizzabile in un contesto reale.

La metodologia fa completo affidamento sull'integrità della piattaforma che interfaccia gli enti privati agli enti richiedenti. Come descritto in III.II, è necessario che il funzionamento della piattaforma sia pienamente trasparente; una politica di distribuzione open source permette all'individuo di sapere nella maniera più dettagliata possibile il funzionamento del meccanismo, incentivando la fiducia nel sistema.

Il sistema di credenziali anonime ha invece lo scopo di disincentivare i contatti al di fuori della piattaforma che mettono a rischio l'integrità del modello distribuito. Basti pensare all'evenienza in cui l'ente richiedente venga a conoscenza dell'ente operativo incaricato della fase di acquisizione e, per motivi politici, corrompa l'ente per indirizzare l'indagine verso risultati che possano favorire la loro posizione pubblica. La credenziale anonima viene fornita ad un ente privato ad ogni partecipazione ad un'indagine. Ovviamente, più enti privati sono attivi sulla piattaforma, più il sistema diventa resistente a problemi di imparzialità, data l'incrementale difficoltà per un ente di scoprire l'identità nascosta degli altri partecipanti all'aumentare del numero degli stessi. Come descritto nella fase di selezione (III.II), il numero totale di partecipanti rimane nascosto a tutti gli enti; questa regola porta delle garanzie a entrambe le parti: l'ente privato sarà disincentivato a richiedere una retribuzione troppo elevata sapendo che il numero di partecipanti è esiguo, e dall'altra parte, l'ente richiedente non potrà giocare al ribasso non sapendo quanti partecipanti rimangono disponibili nel caso in cui la fase di contrattazione non vada a buon fine.

Durante la fase di discussione gli enti operativi non sanno ancora la fase di iter alla quale verranno assegnati, questa regola stimola gli enti operativi a fornire contributi per la pianificazione dell'indagine su ogni fronte della stessa. Come citato nella fase di contrattazione, nel caso in cui l'iterazione attuale fallisca, tutti i commenti effettuati nella fase di discussione vengono resi disponibili agli enti operativi eletti per l'iterazione successiva; questo modus operandi permette di snellire la procedura riutilizzando materiale potenzialmente valido.

Per quanto concerne la fase di contrattazione, all'ente richiedente è fatto divieto di visualizzare le retribuzioni richieste da ogni singolo ente operativo per non creare falle nel sistema di credenziali anonime. Un esempio di exploit del meccanismo sarebbe quello codificare nell'importo del singolo l'identità di un ente operativo, con un protocollo di riconoscimento stabilito a priori. Le regole che stabiliscono le modalità di proposta di remunerazione sono state pensate per evitare movimenti speculativi tra le parti.

La fase di esecuzione comprende nozioni che regolano l'interazione tra le parti durante il nucleo dell'indagine. Sono state volontariamente evitate regolamentazioni inerenti a

tecnicismi e flussi di lavoro per promuovere l'auto-organizzazione degli enti operanti e perché risulterebbe impossibile definire regole specifiche rispetto alla pletora di scenari d'indagine differenti. Per quanto concerne i casi di discordanza durante le revisioni eseguite dagli enti controllori, si è scelto di adottare un *modus operandi* che richiede l'unanimità dei voti. Se da un lato la procedura risulta essere macchinosa, dall'altro lato viene permesso ad ogni ente di esprimere solleciti tecnici e osservazioni che passerebbero inosservate in un sistema di maggioranza. Il caso estremo di continuo rigetto - sia in minoranza che in maggioranza - e le rispettive regole, sono state create per ottenere un metodo senza situazioni di stallo, si noti che nel momento in cui tutti gli enti privati sono composti da personale con una formazione di alto livello, queste eventualità sono destinate ad insorgere raramente.

Collegandosi al paragrafo precedente, è importante che l'abilità degli enti privati di condurre indagini sia in qualche modo certificata. Sotto il profilo soggettivo, agli enti privati che parteciperanno all'indagine verranno richiesti requisiti minimi di competenza che dovranno essere certificati.

Infine, alcune considerazioni sui due documenti che compongono la metodologia. Per quanto riguarda il Documento di *Fundamentum* (III.II), viene imposto l'obbligo di non inserire alcuna nota di tipo tecnico per evitare instradamenti iniziali (dolosi o colposi) che distorcerebbero il naturale sviluppo dell'indagine da parte degli enti operanti. Il Documento di *Relatio* (III.III) invece risulta apportare diversi vantaggi alla metodologia:

- Viene promossa la tracciabilità di ogni apporto all'indagine.
- Viene promossa la qualità della documentazione.
- Facilita il team work remoto.
- Facilita il riutilizzo di dati in possibili indagini future.

#### **IV.II. Criticità sulla metodologia.**

La metodologia proposta ha come principale obiettivo la minimizzazione di *bias* delle indagini. Le regole proposte in III.I sono state difatti pensate e formulate sulla base di questo obiettivo. Conseguentemente, il modello ha diversi svantaggi che vengono discussi in questa sezione.

Il primo grande svantaggio concerne i costi e i tempi di esecuzione. L'implementazione della filosofia *Distributed Ledgers* comporta l'aumento dei tempi totali dell'indagine. I costi aumentano non solo per il numero di enti coinvolti ma anche per la quantità di lavoro necessaria ad attuare il modello: ogni ente operativo riveste il ruolo di ente esecutivo una volta e il ruolo di ente controllore cinque volte. I problemi dovuti ai tempi di esecuzione sorgono solo quando le decisioni governative collegate all'indagine devono essere adottate in tempi stretti; a titolo di esempio basti pensare all'adozione delle misure di contenimento durante l'emergenza Covid-19. Negli altri casi, ossia in quelli dove non vi è un'urgenza decisionale, si confida sul fatto che l'aumento dei tempi di indagine è pienamente giustificabile al fine di ottenere risultati più affidabili. Per quanto riguarda lo svantaggio del costo, vogliamo portare come esempio a supporto dell'adozione dell'MDA due indagini estere. La prima riguarda un sondaggio effettuato nel Regno Unito<sup>[79]</sup>, dove viene stimato che "il 53% del settore pubblico e di organizzazioni no-profit non sanno quantificare quanto budget viene sperperato [ogni anno] a causa dell'inaccuratezza dei dati"<sup>[80]</sup>. Sempre dallo

stesso sondaggio riportiamo “Il 91% delle organizzazioni ha ammesso che del budget è stato sprecato come risultato di bassa qualità del dato. Di media, il 12% del budget dipartimentale [annuale o quadrimestrale], viene sprecato - una situazione che è comune a tutti i settori intervistati”<sup>[81]</sup>. La seconda indagine riguarda un sondaggio effettuato in Nuova Zelanda<sup>[82]</sup> dove hanno partecipato 29 enti (tra cui organizzazioni governative, banche e imprese), ove viene affermato che “la bassa qualità del dato può originare di media costi con un significativo impatto del 10% delle entrate annuali”<sup>[83]</sup>. Ulteriori studi sono richiesti nel territorio italiano per stimare i benefici derivanti da un incremento della qualità dei dati prodotti; ciononostante la visione in cui *indagini costose ma di alta qualità portano a benefici sul lungo termine* non sembra apparire affatto come un’illusione.

Un altro grande svantaggio può attenerne alla compatibilità del modello con la formazione dei dati aggregati personali, in quanto possono sorgere conflitti con la normativa prevista a loro protezione (GDPR). Al riguardo si ritiene però che siano maturati i tempi per un cambio di prospettiva: non si possono continuare ad applicare le logiche sulla protezione dei dati personali, logiche nate per proteggere i dati degli individui dalle *big five*<sup>[84]</sup> che perseguono il profitto, anche per le pubbliche amministrazioni il cui fine è dettato dall’art. 97 Cost: imparzialità e buon andamento per il perseguimento dell’interesse pubblico e della collettività<sup>[85]</sup>. La procedura è strutturata per garantire la sindacabilità, da parte del cittadino, delle scelte adottate dall’amministrazione: l’ottica è di cooperazione fra privato cittadino e amministrazione pubblica affinché quest’ultima possa adottare decisioni più trasparenti, sindacabili e corrette. Inoltre, mentre la formazione dei *big data* da parte di pochi *big players* può ledere la privacy<sup>[86]</sup> da parte dell’amministrazione questo problema non si pone perché l’iter di formazione dei dati sarà preceduto da una dichiarazione di intenti (ref. III.III) che specificherà previamente a quale scopo<sup>[87]</sup> i dati verranno formati nel pieno rispetto del principio di finalizzazione dell’amministrazione pubblica agli interessi pubblici<sup>[88]</sup>.

Infine, le fasi sono state pensate in modo tale da fornire un’applicazione ottimale del modello proposto. La proposta è fatta nella consapevolezza che vi sono norme che stabiliscono vincoli legali per appalti di un determinato valore (appalti sopra e sotto-soglia). Dato il carattere innovativo e il grado di complessità della proposta, se essa non fosse realizzabile secondo quanto esposto di seguito, si auspica, in via sussidiaria, l’utilizzo di quegli strumenti che l’AgId (Agenzia per l’Italia digitale) definisce “*appalti pubblici di soluzioni innovative*” e che sono contemplati all’interno del Codice dei contratti pubblici (art. 59 e seg. del d.lgs. del 18 aprile 2016, n. 50) quali la procedura competitiva con negoziazione (art. 62), la procedura negoziata senza previa pubblicazione di un bando di gara (art. 63), il dialogo competitivo (art. 64), il partenariato per l’innovazione (art. 65)<sup>[89]</sup>.

## **V. Conclusione.**

Alla luce delle considerazioni esposte in questa sede, è possibile esprimere qualche contrappunto conclusivo. In generale, si è potuto constatare che non vi è normativa o regolazione puntuale né sui *big data* né in riferimento alla loro formazione né per quanto riguarda le fasi successive che culminano nella decisione politica/amministrativa. Al contempo, una regolazione è però necessaria in quanto l’UE si sta muovendo nella direzione di uno sfruttamento massivo dei dati per incrementare la crescita economica. Ciò nondimeno, la legislazione e l’apparato amministrativo dimostrano qualche difficoltà ad

approntare un quadro normativo coerente con i nuovi obiettivi sovranazionali, senza contare che, ad oggi, non vi sono politiche di trasparenza sull'origine e sulla formazione degli stessi. È chiaro che l'inevitabile fenomeno *de quo* rappresenta un importante strumento in mano alle amministrazioni che sono già in possesso di una quantità smisurata di dati: la proposta rappresenta una linea-guida, un metodo, che può permettere alle amministrazioni di formare e gestire nuovi dati secondo un modello improntato alla trasparenza e alla sindacabilità delle decisioni politiche e amministrative.

Da qui l'esigenza di ipotizzare e, eventualmente, proporre un procedimento alternativo per la formazione dei dati ispirato alla filosofia dei *Distributed Ledgers* che, coinvolgendo i privati nell'iter di formazione dei dati, attua una forma di coordinamento e collaborazione fra questi e la pubblica amministrazione. In particolare, il Modello Decentralizzato Anonimo (ref. IV.I) è idoneo a scongiurare la gran parte dei rischi che possono manifestarsi nel processo che dai dati conduce alla conoscenza quali la scarsa qualità, usi impropri e/o manipolatori, errori e *bias*, ottenendo così dati il più possibile genuini, documentati e integri. Gli svantaggi derivanti dall'adozione di tale modello (ref. IV.II) quali l'aumento dei costi e dei tempi d'indagine, la difficile applicazione ai dati personali, il rispetto della normativa di evidenza pubblica, sono d'altra parte risolvibili soprattutto alla luce del fatto che il fine ultimo è quello di garantire al cittadino l'accesso ad informazioni veritiere scevre da condizionamenti e manipolazioni.

1. Luxembourg Centre for Systems Biomedicine - University of Luxembourg, Università degli Studi di Milano - Dipartimento di Informatica. [↑](#)
2. Università del Piemonte Orientale. [↑](#)
3. Per bibliografia sui rapporti tra diritto e tecnologia si veda il pionieristico studio di Frosini V. (1968), *Cibernetica, diritto e società*, Milano, Edizioni Comunità; riferimento essenziale rimangono, tra gli altri, Losano M. G. (1969), *Giuscibernetica. Macchine e modelli cibernetici nel diritto*, Torino, Einaudi; Frosini V. (1981), *Il diritto nella società tecnologica*, Milano, Giuffrè; Di recente, v. Faini F., Pietropaoli S. (2017), *Scienza giuridica e tecnologie informatiche*, Torino, Giappichelli; Falcone M. (2017), "Big data" e pubbliche amministrazioni: nuove prospettive per la funzione conoscitiva pubblica, in *Riv. trim. dir. pubb.*, n. 3/2017, pp. 601-639; Taddei Elmi G. (2014), *Informatica e Diritto. Presupposti, storia, disciplina, insegnamento, ius condendum*, in *Informatica e diritto*, n. 2/2014, pp. 85-123; Ziccardi G., Perri P., *Tecnologia e diritto. Fondamenti d'informatica per il giurista*, Milano, Giuffrè, 2017.  
Più nello specifico, sui rapporti tra diritto e *data analysis* si veda Forte. P. (2020), *Diritto amministrativo e data science. Appunti di intelligenza amministrativa artificiale (AAI)*, in *PA persona e amministrazione*, n. 1/2020, pp. 247-313; Cavallo Perin R. (2021), *Pubblica amministrazione e data analysis*, in Cavallo Perin R. (a cura di), *L'amministrazione pubblica con i big data: da Torino un dibattito sull'intelligenza artificiale*, Università degli studi di Torino, pp. 11-18; Ponti B. (2021), *L'amministrazione come fornitore e come fruitore di dati personali pubblici: sono praticabili soluzioni basate sulla Big Data Analytics/Machine Learning*, *ibidem*, pp. 183-190. [↑](#)
4. Cfr. Faro S. (2020), *Prospettive di sviluppo dell'informatica giuridica*, in Faro S., Frosini T. E., Peruginelli G., (a cura di), *Dati e algoritmi. Diritto e diritti nella società digitale*, Bologna, il Mulino, p. 71: "Se il diritto si caratterizza come ordinamento, cioè fenomeno di organizzazione della comunità fondata sugli interessi e i valori in essa diffusi, allora la produzione, l'interpretazione e l'applicazione delle norme giuridiche destinate a regolare la realtà sociale non possono rimanere indifferenti ad acquisizioni scientifiche e metodologiche in grado di illuminare le dinamiche sociali e i processi individuali e collettivi a esse sottesi". Per una definizione di dato v. Pascuzzi G. (2020), *Il diritto dell'era digitale*, quinta edizione, Bologna, Il Mulino, pp. 36-37. [↑](#)
5. Cfr. fra i tanti, De Leonardis F. (2021), *Il principio di precauzione nella gestione dell'emergenza epidemiologica*, in Malvicini M. (a cura di), *Il governo dell'emergenza. Politica, scienza e diritto al cospetto della pandemia COVID-19*, pp. 51-76; Piazza A. (2021), *Pandemia e contagio dei numeri*, in

- Pajno A., Violante L. (a cura di), *Biopolitica, pandemia e democrazia. Rule of law nella società digitale*, Vol. II Bologna, il Mulino, pp. 89-106; Pajno A. (2021), *Democrazia e governo della pandemia*, in Pajno A., Violante L. (a cura di), *Biopolitica, pandemia e democrazia. Rule of law nella società digitale*, Vol. I, Bologna, il Mulino, pp. 31-65. [↑](#)
6. Zeno-Zencovich (2018), *Dati, grandi dati, dati granulari e la nuova epistemologia del giurista*, in *mediaLaws*, n. 2/2018, pp. 36-37: “è necessario comprendere quali dati e quali metodiche sono stati utilizzati. È noto da quasi un secolo che le statistiche subiscono la influenza soggettiva di chi le raccoglie ed elabora. Lo stesso può dirsi della analisi dei dati i quali, ancora più facilmente, possono essere piegati per giustificare talune soluzioni”. [↑](#)
  7. Per riflessioni attinenti al tema della qualità dei dati v. Carloni E. (2021), *Qualità dei dati, big data e amministrazione pubblica*, in Cavallo Perin R. (a cura di), *L'amministrazione pubblica con i big data: da Torino un dibattito sull'intelligenza artificiale*, Università degli studi di Torino, pp. 117-130. L'A. a pag. 120 afferma che le p.a. debbono porre attenzione ai procedimenti di formazione dei dati al fine di utilizzare dati che siano di qualità. [↑](#)
  8. Galetta D.U. (2016), *La trasparenza, per un nuovo rapporto tra cittadino e pubblica amministrazione: un'analisi storico-evolutiva, in una prospettiva di diritto comparato ed europeo*, in *Rivista Italiana di Diritto Pubblico Comunitario*, n. 5/2016, pp. 1019-1065; Sassi S. (2019), *Gli algoritmi nelle decisioni pubbliche tra trasparenza e responsabilità*, in *Analisi Giuridica dell'Economia*, n. 1/2019, pp. 109-128. [↑](#)
  9. Si preferisce far riferimento al concetto di modernizzazione e non di progresso aderendo così all'impostazione di Sciarrone R. (2002), *Voce: Modernizzazione*, in *Nova l'Enciclopedia UTET*, Torino, UTET, pp. 99-101. [↑](#)
  10. Il progresso tecnologico si divide in quattro principali aree di sviluppo: banda larga e telecomunicazioni, economica Internet, politica dei consumatori, stato del digitale, in Floridi L. (2020), *Il verde e il blu*, Milano, Raffaello Cortina Editore, p. 94. In tema di *governance* si segnala la “[proposta di regolamento del parlamento europeo e del consiglio relativo alla governance europea dei dati \(atto sulla governance dei dati\)](#)”. [↑](#)
  11. Si fa riferimento al periodo della “Nuova agenda strategica 2019-2024” del Consiglio Europeo. [↑](#)
  12. Nelle prime comunicazioni adottate dall'UE sul tema, i termini dati, documenti e informazioni vengono utilizzati in modo intercambiabile. [↑](#)
  13. Dato ed informazione sono concetti distinti seppur interdipendenti. Cfr. Rossi dal Pozzo F. (2020), *Il mercato unico digitale europeo e il regolamento ue sulla privacy*, in Cavallo Perin R. e Galetta D.U. (a cura di), *Il diritto dell'amministrazione pubblica digitale*, Torino, Giappichelli, pp. 43-44; Carullo G. (2020), *Dati, banche dati, blockchain e interoperabilità dei sistemi informatici nel settore pubblico*, in *ibidem*, pp. 192-194; Rossa S. (2020), *Trasparenza e accesso all'epoca dell'amministrazione digitale*, in *ibidem*, pp. 269-270. [↑](#)
  14. La definizione di riutilizzo è contenuta all'art. 2 del d.l. 36/2006. [↑](#)
  15. Le misure di finanziamento e di sostegno sono elencate al punto 5.3 della comunicazione. [↑](#)
  16. Nel dicembre 2012 veniva lanciato il portale [open data dell'UE](#). Dal 2011 anche in Italia vi è un portale che raccoglie [i dati aperti della pubblica amministrazione](#) che dal 2015 è gestito dall'[Agenzia per l'Italia Digitale](#). Il portale è stato creato in virtù dell'art. 9 d.lgs. n. 36/2006 così come modificato dal d.lgs. n. 102/2015 in attuazione della direttiva 2003/98/CE relativa al riutilizzo di documenti nel settore pubblico. Anche la [regione Piemonte](#) ha un portale per gli open data con 948 dataset disponibili, essa è stata la prima regione ad aver creato un portale per gli open data, e ha legiferato in materia un anno prima rispetto allo Stato, cfr. Faini F. (2019), *Data society - Governo dei dati e tutela dei diritti nell'era digitale*, Milano, Giuffrè Francis Lefebvre, p. 155. L'iniziativa piemontese è altresì citata nella comunicazione della commissione europea del 2011 ([COM\(2011\) 882](#)) p. 5. [↑](#)
  17. Cfr. Falcone M. (2021), *La funzione conoscitiva nella rivoluzione dei dati*, in Cavallo Perin R. (a cura di), *L'amministrazione pubblica con i big data*, cit., pp. 183-190; Rossa S. (2020), *Trasparenza e accesso all'epoca dell'amministrazione digitale*, cit., p. 207; Faini F. (2019), *Data society*, cit. p. 62. [↑](#)
  18. Direttiva 2013/37/UE considerando n. 4 e Direttiva 2019/1024/UE considerando n. 14 [↑](#)
  19. La definizione di big data è fortemente dibattuta in letteratura; in termini ampi, è possibile far riferimento a “*grandi quantità di dati, di natura diversa e processati ad elevata velocità*” (Laney D. (2001), *3D Data Management: Controlling Data Volume, Velocity, and Variety*). Per la Commissione europea, il termine “big data” fa riferimento a grandi quantità di dati di tipo diverso prodotti a grande velocità da numerosi tipi di fonti. La gestione di questi *dataset* ad elevata variabilità e in

tempo reale impone il ricorso a nuovi strumenti e metodi, quali ad esempio potenti processori, *software* e algoritmi, (COM (2014) 442 final).

In generale, le tre caratteristiche fondamentali dei “big data” sono: volume, velocità e varietà, integrate da alcuni autori con la veridicità e il valore. Sulle caratteristiche dei “big data” v. Giardullo P. (2020), *Algoritmi, dati e piattaforme*, in Magaudo P. e Neresini F., *Gli studi sociali sulla scienza e la tecnologia*, Bologna, il Mulino, p. 213; Paccagnella L. (2020), *Sociologia della comunicazione nell’era digitale*, Bologna, il Mulino, p. 199; Bennato v. (2015), *Il computer come macroscopio. Big data e approccio computazionale per comprendere i cambiamenti sociali e culturali*, Milano, FrancoAngeli. Cfr. letteratura straniera: [Brynjolfsson E. e McAfee A., \*Big data’s management revolution. The promise and Challenge of Big Data\*, in Harvard Business Review Insight Center Report, 11 settembre 2012;](#) [Demchenko Y., Ngo C. e Membrey P., \*Architecture framework and components for the big data ecosystem\*, in Journal of System and Network Engineering, 12 settembre 2013.](#) Il tema è trattato diffusamente in Faini F. (2019), *Data society*, cit., pp. 160-173.

La definizione di *big data* non va confusa con quella di *open data*, cfr. Carullo G. (2020), *Dati, banche dati, blockchain e interoperabilità dei sistemi informatici nel settore pubblico*, cit., p. 198. Anche se, al fine di ridurre l’asimmetria di potere che caratterizza il mondo dei *Big Data* si auspica che essi possano essere rilasciati come *open data*, sul punto v. Pietropaoli S. (2020), *Habeas Data. I diritti umani alla prova dei Big Data*, in Faro S., Frosini T., Peruginelli G. (a cura di), *Dati e algoritmi*, Bologna, il Mulino, p. 111. [↑](#)

20. Cfr. anche Sola A. (2020), *Utilizzo dei big data nelle decisioni pubbliche tra innovazione e tutela della privacy*, in *mediaLaws* n. 3/2020, pp. 202-203. [↑](#)
21. Per una definizione di interoperabilità v. Carullo G. (2021), *Interoperabilità dei dati e riflessi organizzativi: il caso della conservazione digitale*, in Cavallo Perin R. (a cura di), *L’amministrazione pubblica con i big data*, cit., pp. 251-259. [↑](#)
22. Le priorità sono le seguenti: economie più forti con più posti di lavoro; società in grado di responsabilizzare e proteggere; un futuro energetico e climatico sicuro; uno spazio sicuro di libertà fondamentali; un’azione congiunta efficace nel mondo. [↑](#)
23. [EUROPEAN UNION, \*EUROPEAN COUNCIL CONCLUSIONS ON THE DIGITAL SINGLE MARKET STRATEGY FOR GROWTH AND EMPLOYMENT\*, 2015, par. 2.](#) [↑](#)
24. Cfr. Rossi dal Pozzo F. (2020), *Il mercato unico digitale europeo e il regolamento UE sulla privacy*, in Cavallo Perin R. e Galetta D.U. (a cura di), *Il diritto dell’amministrazione pubblica digitale*, Torino, Giappichelli, p. 47. [↑](#)
25. Sull’interoperabilità dei sistemi informatici si rimanda alle riflessioni contenute in Carullo G. (2020), *Dati, banche dati, blockchain e interoperabilità dei sistemi informatici nel settore pubblico*, in *ibidem*, pp. 209-212. [↑](#)
26. In quest’ottica, è senz’altro di interesse che per la prima volta dal 2003 l’UE ribadisca l’impegno affinché i dati utilizzati siano di qualità. Invero, abbiamo già visto un duplice riferimento alla qualità del dato: la prima volta nella direttiva del 2013 in termini di benefici derivanti dal riutilizzo del dato e la seconda nella comunicazione del 2014 ove si afferma che per pianificare una economia basata sui dati è necessario utilizzare *dataset* che siano affidabili e di qualità. Il riferimento alla “qualità del dato”, pertanto, sia nel 2013 che nel 2014, non è stato fatto in termini di politiche pubbliche o di impegni assunti dall’UE per implementarla. [↑](#)
27. In particolare, con il Regolamento si è stabilito che la libera circolazione dei dati personali nell’Unione non può essere limitata né vietata per motivi attinenti alla protezione delle persone fisiche con riguardo al trattamento dei dati personali; determinandone un incentivo alla circolazione. Con la comunicazione, parimenti, si stigmatizzano le restrizioni ingiustificate alla circolazione dei dati incentivandone il libero flusso “in tutta la catena di valore per fini scientifici, sociali e industriali”. [↑](#)
28. In questa, si afferma che per incrementare la crescita economica è necessario “mettere a disposizione più dati”. [↑](#)
29. Si evidenzia che il Regolamento (UE) 2016/679 ha ad oggetto invece la circolazione dei dati personali. Questi due regolamenti, come evidenziato in Faini F. (2020), *La dimensione giuridica dei Big Data*, in Faro S., Frosini T., Peruginelli G. (a cura di), *Dati e algoritmi*, Bologna, il Mulino, pp. 116-120, costituiscono il *framework* giuridico di riferimento per la circolazione dei dati nell’UE. [↑](#)
30. V. anche Pascuzzi G. (2020), *Il diritto dell’era digitale*, cit., p. 271. [↑](#)
31. Cfr. anche Rossi dal Pozzo F. (2020), *Il mercato unico digitale europeo e il regolamento UE sulla privacy*, cit., pp. 47-48; Carullo G. (2020), *Dati, banche dati, blockchain e interoperabilità dei sistemi*

*informatici nel settore pubblico*, cit., p. 192. [↑](#)

32. Altrettanto importanti sono gli approcci B2B E B2G che prevedono rispettivamente la condivisione e l'utilizzo dei dati da privati da parte di altre imprese e da privati da parte delle autorità pubbliche. [↑](#)
33. Cfr. Rossi dal Pozzo F. (2020), *Il mercato unico digitale europeo e il regolamento UE sulla privacy*, cit., pp. 47-48. [↑](#)
34. Sui profili di responsabilità per la qualità dei dati e democrazia v. Zeno-Zencovich V. (2020), *Big data e epistemologia giuridica*, in Faro S., Frosini T., Peruginelli G. (a cura di), *Dati e algoritmi, Diritto e diritti nella società digitale*, Bologna, il Mulino, pp. 23-24. [↑](#)
35. Il riferimento è alle nuove tecnologie digitali basate sui registri distribuiti (*Distributed Ledgers* di cui si vedrà meglio *infra*). [↑](#)
36. Pascuzzi G. (2020), *Il diritto dell'era digitale*, cit., p. 269. [↑](#)
37. Sulla correlazione fra dati e trasparenza vi riflette in modo illuminante Rossa S. (2020), *Trasparenza e accesso all'epoca dell'amministrazione digitale*, cit., pp. 269-272, per il quale il binomio "accesso e riuso" del dato è fondamentale per rendere ancora più effettivo il principio di trasparenza. Cfr. più diffusamente Faini F. (2019), *Data Society*, cit., pp. 72-116. [↑](#)
38. Al giorno d'oggi gran parte delle analisi effettuate in processi decisionali pubblici sono basate su *big data* poiché all'aumentare del volume di dati aumenta anche l'affidabilità dei risultati prodotti dall'analisi.  
In questo senso v. anche l'[indagine conoscitiva sui Big Data](#) p. 71, nella quale si afferma che "la raccolta e l'analisi di una quantità sempre maggiore di dati porta a una conoscenza più approfondita dei processi e dei clienti e consente di adottare decisioni in grado di migliorare ogni aspetto dell'attività di impresa".  
In senso contrario v. Paccagnella L. (2020), *Sociologia della comunicazione nell'era digitale*, Bologna, il Mulino, pp. 200-201. Sul fatto che dall'elaborazione di grandi numeri non derivi, giocoforza, maggiore oggettività o accuratezza l'A. cita anche [boyd d. e Crawford K. \(2012\), Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon, in Information, Communication & Society, Vol. 15, pp- 662-679.](#)  
Sul fatto che qualità e quantità sono due caratteristiche del dato inscindibili e che all'aumentare della seconda debba aumentare anche la prima v. Giovannini E. (2017), *Scegliere il futuro. Conoscenza e politica al tempo dei big data*, Bologna, il Mulino, p. 132. [↑](#)
39. Il CAD è stato sottoposto a numerosi interventi correttivi esclusi quelli del 2020 ad opera dei decreti cd. rilancio e semplificazione, ciò ad evidenziare quanto sia rapido il progresso tecnologico (cfr. Mangaro F., *Evoluzione ed involuzione delle discipline normative sull'accesso a dati, informazioni ed atti delle pubbliche amministrazioni*, in *Diritto Amministrativo*, fasc. 4, 1° dicembre 2019, pag. 743). Come evidenzia efficacemente Rossa S. (2019), *Open data e amministrazioni regionali e locali. Riflessioni sul processo di digitalizzazione partendo dall'esperienza della regione Piemonte*, in *Diritto dell'Informazione e dell'Informatica (II)*, fasc. 4, pag. 1121: "è vero che il CAD stabilisce dei principi cardine, ma essi possono essere concretizzati unicamente sul piano regolamentare, in conseguenza della repentina e dinamica evoluzione tecnologica che non può essere disciplinata efficacemente da strumenti normativi quali gli atti aventi forza di legge, astratti, generali e caratterizzati da una lenta approvazione con tutti i rischi che ciò comporta". [↑](#)
40. Per l'analisi della norma si rinvia a Rossa S. (2019), *Open data e amministrazioni regionali e locali. Riflessioni sul processo di digitalizzazione partendo dall'esperienza della regione Piemonte*, cit., pag. 1121; Coronidi F. (2018), *Commento art. 50*, in C. Boccia, C. Contessa, E. De Giovanni (a cura di), *Codice dell'amministrazione digitale*, Piacenza, La Tribuna, p. 214. [↑](#)
41. La stessa disposizione è altresì contenuta all'art. 2 c. 1 lett. i) del d.lgs. 36/2006 attuativo della direttiva 2003/98/CE relativa al riutilizzo dei dati nel settore pubblico. Un ulteriore riferimento alla formazione del dato era contenuto all'art. 1 c. 1 lett. m) che definiva il dato delle pubbliche amministrazioni come quel dato formato o comunque trattato da una pubblica amministrazione. La disposizione, seppur sia stata abrogata dal d.lgs. 179/2016, è utile per comprendere come il procedimento di formazione sia qualcosa di separato e diverso dalla trattazione dei dati e da qualsiasi altra attività a loro riconducibile. [↑](#)
42. Art. 1 c. 1 lett. p) e p-bis) del CAD. [↑](#)
43. Se invece accedessimo alla tesi per la quale dato e documento rappresentano, giuridicamente, il medesimo concetto allora questo ragionamento risulterebbe superfluo. In tal senso v. [Minazzi F. \(2013\), Il principio dell'open data by default nel codice dell'amministrazione digitale: profili](#)

- [interpretativi questioni metodologiche, in \*federalismi.it\*, n. 23/2013, p. 4](#). L'A. si riferisce alla definizione, identica, di titolare del dato contenuta nel d.lgs. n. 36/2006 art. 2 lett. i) poiché non era ancora stata inserita nel CAD. [↑](#)
44. Cfr. Giardullo P. (2020), *Algoritmi, dati e piattaforme*, cit. p. 214 [↑](#)
45. I risultati intermedi sono stati pubblicati in un primo report: [Big data Interim report nell'ambito dell'indagine conoscitiva di cui alla delibera n. 217/17/CONS](#); e quelli finali sono stati diffusi il 10 febbraio 2020: [Indagine conoscitiva sui Big Data](#). [↑](#)
46. *Ibidem*, pag. 8. Parla di filiera pubblica dei *Big Data* con particolare attenzione ai dati personali Sola A. (2020), *Utilizzo dei big data nelle decisioni pubbliche*, cit., pp. 211-217.  
Nello scritto si propone una separazione in fasi alternativa, concettualmente identica a quella prospettata dalle autorità garanti ma strutturata in modo da permettere una più compiuta comprensione del modello decentralizzato proposto (Vedi *infra* III.I). [↑](#)
47. Definito anche come “*Data driven decision making*”; indagine conoscitiva, pp. 8-21. [↑](#)
48. Sul principio di trasparenza cfr. Faini F. (2019), *Data society*, cit. pp. 74-83; v. anche Rossa S. (2020), *Trasparenza e accesso all'epoca dell'amministrazione digitale*, cit., pp. 247-275. [↑](#)
49. Indagine conoscitiva, pp. 65-66. [↑](#)
50. Cfr. Faro S. (2020), *Prospettive di sviluppo dell'informatica giuridica fra Big Data e scienze sociali computazionali*, in Faro S., Frosini T. E., Peruginelli G. (a cura di), *Dati e algoritmi*, Bologna, il Mulino, pp. 69-70, sul ruolo delle simulazioni nell'attività di *policy making*. [↑](#)
51. Mantelero A. (2012), [Big data: i rischi della concentrazione del potere informativo digitale e gli strumenti di controllo](#), in *Il diritto dell'informazione e dell'informatica*, fasc. 1, p. 135; Pascuzzi G. (2020), *Il diritto dell'era digitale*, cit., p. 344; Merloni F. (2008), *Trasparenza delle istituzioni e principio democratico*, in Merloni F., Arena G., Corso G., Gardini G., Marzuoli C., (a cura di), con la collaborazione di Carloni E., *La trasparenza amministrativa*, Milano, Giuffrè, pp. 3-27; Orefice M. (2016), *I big data. Regole e concorrenza*, in *Politica del diritto*, n. 4/2016, pp. 697-743; Ziccardi G. (2015), *Internet, controllo e libertà. Trasparenza, sorveglianza e segreto nell'era tecnologica*, Milano, Raffaello Cortina Editore; Cuniberti M. (2019), *Potere e libertà nella rete*, in Manetti M. e Borrello R. (a cura di), *Il diritto dell'informazione*, Modena, Mucchi Editore, p. 44; Floridi L. (2020), *Il verde e il blu*, cit., pp. 195-204, ove si possono leggere accurate suggestioni su quello che l'A. definisce “*il potere grigio digitale*”. [↑](#)
52. L'indagine rileva, in termini generali, aspetti critici riguardo ad approcci centralizzati. Due sono i casi emblematici: l'uno è l'ISTAT e l'altro è uno strumento di recentissima attuazione previsto (2021) all'art. 50 *ter* del CAD che implementando il Data & Analytics Framework (DAF) ha costituito la Piattaforma Digitale Nazionale Dati (PDND). In particolare, si è rilevato come la creazione di tale piattaforma comporti “un accentramento e una duplicazione di tutti i dati detenuti dalle pubbliche amministrazioni per finalità del tutto generiche, realizzando di fatto una concentrazione presso un unico soggetto di informazioni, anche sensibili e sensibilissime, con evidenti rischi di vulnerabilità dei dati stessi ovvero di possibili usi distorti”.  
L'art. 50 *ter* CAD assegna un ruolo centrale alla Presidenza del Consiglio dei ministri conferendogli compiti di progettazione, sviluppo e realizzazione della PDND e, come notato dalle autorità garanti nell'indagine, ha finalità istituzionali molto generali: favorire la conoscenza, l'utilizzo e la condivisione delle informazioni e dei dati. [↑](#)
53. Cfr. in tema di piattaforme online con riferimento ai dati detenuti dalle pubbliche amministrazioni presso i *data center*: v. Carullo G. (2020), *Dati, banche dati, blockchain e interoperabilità dei sistemi informatici nel settore pubblico*, cit., p. 201. [↑](#)
54. Indagine conoscitiva, p. 96. [↑](#)
55. Trattando di conservazione e gestione dei dati tramite sistemi informatici decentralizzati può essere suggestivo far riferimento alla nuova tecnologia a registro distribuito c.d. *blockchain* che comporta una condivisione dei dati con i privati o con altre pubbliche amministrazioni che ne divengono parimenti detentori. È palese che questo tipo di tecnologie a registro distribuito (*Distributed Ledger Technologies*, DLT) consentano un maggior rispetto del principio di trasparenza grazie alla condivisione dei dati fra privati e pubbliche amministrazioni. Cfr. Carullo G. (2020), *Dati, banche dati, blockchain e interoperabilità dei sistemi informatici nel settore pubblico*, cit., pp. 202-205: “*un sistema basato su DLT permetterebbe ai cittadini di essere essi stessi co-depositari delle informazioni di loro interesse, potendovi così accedere direttamente senza l'intermediazione di servizi volti a consentire l'accesso ai dati*”. [↑](#)

56. Carloni E. (2014), *L'amministrazione aperta. Regole strumenti limiti dell'open government*, Santarcangelo di Romagna, Maggioli; Faini F. (2013), *La strada maestra dell'open government: presupposti, obiettivi, strumenti*, in *Cyberspazio e diritto*, n. 2/2013, pp. 213-238; Di Donato F. (2010), *Lo stato trasparente. Linked open data e cittadinanza attiva*, Pisa, Edizioni ETS. [↑](#)
57. Pascuzzi G. (2020), *Il diritto dell'era digitale*, cit., p. 274. Cfr. Faini F. (2019), *Data society*, cit., pp. 15-17, secondo l'A. (alla quale si rimanda anche per ulteriore bibliografia sull'*open government*) le amministrazioni *open government* sono quelle in grado di garantire la trasparenza a tutti i livelli, rendendo le attività e i dati dell'amministrazione aperte e disponibili. Sull'*open government* si veda altresì Galetta D.U. (2019), *Open-Government, open-data e Azione Amministrativa*, in *Istituzioni del federalismo*, n. 3/2019, pp. 663-683. [↑](#)
58. Cfr. Faini F. (2019), *Data society*, cit., p. 17 che sul punto cita anche Sartori L. (2013), *Open government: what else?*, in *Istituzioni del federalismo*, fasc. 3-4, pp. 753-775. [↑](#)
59. Così come previsto dall'art. 1 del d.lgs. 33/2013. [↑](#)
60. Faini F. (2019), *Data society*, cit., p. 18. V. anche Rossa S. (2020), *Trasparenza e accesso all'epoca dell'amministrazione digitale*, cit., p. 249, per il quale la strategia *Open Government* realizza un "modello di amministrazione finalizzato a porre in essere una relazione più salda e stretta fra il settore pubblico, Pubblica Amministrazione in primis, e gli altri attori privati, in particolare i cittadini". [↑](#)
61. Cfr. Rossa S. (2020), *Trasparenza e accesso all'epoca dell'amministrazione digitale*, cit., pp. 269-270; indagine conoscitiva, p. 270. [↑](#)
62. Sul rapporto funzionale tra digitalizzazione e processo di inclusione e partecipazione democratica v. Fasano G. (2019), *Le decisioni automatizzate nella pubblica amministrazione: tra esigenze di semplificazione e trasparenza algoritmica*, in *mediaLaws*, n. 3/2019, pp. 234-241. [↑](#)
63. Faini F. (2019), *Data society*, cit., pp. 177-178. [↑](#)
64. Di Nunzio G. M. (2015), *Basi di dati: Manuale di esercizi per la progettazione concettuale*, Bologna, Esculapio Editore. [↑](#)
65. Indagine conoscitiva sui *big data*, pp. 8-22 [↑](#)
66. Con il termine inglese *bias* si intende un qualsiasi effetto, indotto volontariamente o non, che distorce la rappresentazione delle misurazioni o osservazioni raccolte, compromettendo quindi l'informazione derivante dal dato. [↑](#)
67. Letteralmente tradotto "grezzo", che non ha subito processi di trasformazione. [↑](#)
68. Corruzione del dato che viene prodotta da interferenze o limiti fisici/tecnologici nella fase di acquisizione. [↑](#)
69. Sia in termini temporali (e.g. diversi dati provengono da acquisizioni effettuate in diversi giorni) che in termini di natura del dato stesso (e.g. lista delle patologie di un individuo e dati anagrafici dello stesso). [↑](#)
70. Verosimilmente la "stazione appaltante" ai sensi dell'art. 3 c. 1 lett. o) del d.lgs. n. 50/2016. [↑](#)
71. Verosimilmente "l'operatore economico" ai sensi dell'art. 3 c. 1 lett. p) del d.lgs. n. 50/2016, che in questa fase del modello sarà un "candidato" ai sensi della lett. bb) dello stesso Codice. Il candidato dovrà essere qualificato dal sistema di *rating* di impresa ai sensi dell'art. 83 comma 10 del Codice e pertanto dovrà possedere la certificazione rilasciata dall'ANAC. [↑](#)
72. Ai sensi dell'art. 213 del d.lgs. n. 50/2016 può essere preposta a questa attività anche l'ANAC (autorità nazionale anticorruzione) che ha funzioni di vigilanza, controllo e regolazione del mercato dei contratti pubblici. [↑](#)
73. [Belenkiy M., Camenisch J., Chase M., Kohlweiss M., Lysyanskaya A., Shacham H. \(2009\), Randomizable Proofs and Delegatable Anonymous Credentials, in Halevi S. \(eds\), Advances in Cryptology - CRYPTO 2009. CRYPTO 2009. Lecture Notes in Computer Science, vol 5677, Berlin, Heidelberg, Springer, pp. 108-125. \[↑\]\(#\)](#)
74. [Camenisch J., Lysyanskaya A. \(2001\), An Efficient System for Non-transferable Anonymous Credentials with Optional Anonymity Revocation, in Pfitzmann B. \(eds\), Advances in Cryptology - EUROCRYPT 2001. EUROCRYPT 2001. Lecture Notes in Computer Science, vol 2045, Berlin, Heidelberg, Springer, pp. 93-118. \[↑\]\(#\)](#)
75. [Kahn M. G., Brown J. S., Chun A. T., Davidson B. N., Meeker D., Ryan P. B., Schilling L. M., Weiskopf N. G., Williams A. E., & Zozus M. N. \(2015\), Transparent reporting of data quality in distributed data networks, Washington, DC, EGEMS, 3\(1\), 1052. \[↑\]\(#\)](#)
76. [Ercole A., Brinck V., George P., Hicks R., Huijben J., Jarrett M., Vassar M., Wilson L., & DAQCOR](#)

- [collaborators \(2020\), \*Guidelines for Data Acquisition, Quality and Curation for Observational Research Designs \(DAQCORD\)\*, \*Journal of clinical and translational science\*, 4\(4\), 354-359.](#) ↑
77. [Cerf M., Matz S., & Berg A. \(2020\), \*Using Blockchain to Improve Decision Making That Benefits the Public Good\*. \*Frontiers in Genetics\*.](#) ↑
78. [Ølnes S., Ubacht J., & Janssen M. \(2017\), \*Blockchain in government: Benefits and implications of distributed ledger technology for information sharing\*, \*Gov. Inf. Q.\*, 34, 355-364.](#) ↑
79. [The Data Advantage: How accuracy creates opportunity - An Experian QAS 2013 Research Report.](#) ↑
80. Citazione originale: "53% of public sector and not-for-profit organisations do not know how much budget is wasted due to data inaccuracies." ↑
81. Citazione originale: "91% of organisations admit that budgets have been wasted over the last 12 months as a result of poor data quality. On average, 12% of departmental budgets have been wasted - a situation that is true across all sectors surveyed." ↑
82. Kahn et al. (2015), *Transparent Reporting of Data Quality in Distributed Data Networks*, cit., pp. 3. ↑
83. Citazione originale: "[...] poor data quality has an average cost impact as high as 10% of an organization's annual revenue" ↑
84. Facebook, Alphabet (Google), Apple, Amazon, Microsoft. ↑
85. Sui rischi di una "dittatura dei dati" da parte dei poteri pubblici v. Faini F. (2019), *Data society*, cit., p. 395. ↑
86. Lalli A. (2019), *Il sapere e la professionalità dell'amministrazione pubblica nell'era dei big data e dell'intelligenza artificiale*, atti del convegno AIPDA, pp. 1-2; cfr. Faini F. (2019), *Data society*, cit., p. 393; Per puntuali riflessioni sul rapporto fra *Big Data* e *privacy* si rimanda a Pietropaoli S. (2020), *Habeas Data. I diritti umani alla prova dei Big Data*, cit., pp. 108-109; Faini F. (2020), *La dimensione giuridica dei Big Data*, cit., p. 123; Faini F. (2019), *Data society*, cit., pp. 358-375. ↑
87. Così come previsto dall'art. 50 c. 2-bis CAD che, nell'interpretazione data da Faini e alla quale si aderisce, evidenzia la necessità che in ambito pubblico l'analisi dei *big data* sia finalizzata alle funzioni istituzionali in modo tale da renderle più efficaci (Faini F. (2019), *Data society*, cit., pp. 179). ↑
88. Casetta E. (2017), *Manuale di diritto amministrativo*, cit., p. 82; Romano A. (2005), *Introduzione*, in Mazarroli L., Pericu G., Romano A., Roversi Monaco F.A., Scoca F.G. (a cura di), *Diritto amministrativo*, Vol. 1, Bologna, Monduzzi. ↑
89. Per notazioni di carattere generale su tali procedure di scelta del contraente v. Cerulli Irelli V. (2017), *Lineamenti del diritto amministrativo*, Giappichelli, Torino, sesta edizione, pp. 524-531; Clarich M. (2019), *Manuale di diritto amministrativo*, il Mulino, Bologna, quarta edizione, pp. 436-443; Casetta E. (2017), *Manuale di diritto amministrativo*, Giuffrè, Milano, diciannovesima edizione, pp. 642-644. ↑