

**#DATIBENECOMUNE**

# ***I DATI CHE VORREI***

***dossier n.1***

***PNRR: quali dati e come pubblicarli***

**a cura di onDATA**

# I DATI CHE VORREI

dossier n.1 – pnrr: quali dati e come pubblicarli

## INDICE

[INDICE](#)

[INTRODUZIONE](#)

[PERCHÉ VOGLIAMO I DATI](#)

[I DATI CHE VORREI](#)

[DATI: COME LI VORREI](#)

[I principi FAIR](#)

[La qualità dei dati](#)

[Riferimenti utili](#)

[ETICA DEI DATI](#)

[CREDITS](#)

[GLOSSARIO](#)

**I dati che vorrei**

a cura dell'[Associazione onDATA](#)

con la collaborazione di ActionAid Italia e Transparency International Italia  
grazie al supporto di Open Society Foundations

Pubblicato nell'ambito della campagna [#datibenecomune](#)

Licenza [CC BY 4.0](#)

Febbraio 2022



## INTRODUZIONE

[#DatiBeneComune](#) è una campagna nata nel novembre 2020 per chiedere al Governo italiano di pubblicare in [formato aperto](#) e accessibile i dati sulla gestione della pandemia di COVID-19.

La campagna è stata accolta sin dall'inizio con entusiasmo da parte di esperti, organizzazioni, testate giornalistiche, che hanno aderito immediatamente all'iniziativa. Ad oggi, oltre 250 organizzazioni si sono ufficialmente unite all'iniziativa e circa 55.000 cittadini e cittadine hanno firmato la petizione connessa.

Da maggio 2021, in collaborazione con l'Osservatorio Civico PNRR, si chiede l'apertura di tutti i dati relativi al [PNRR](#) (Piano Nazionale di Ripresa e Resilienza) e, più in generale, l'apertura di tutti i dati di pubblico interesse.

"I dati che vorrei", grazie all'intervento di autorevoli esperti ed esperte di diverse realtà della società civile, intende delineare le caratteristiche e le modalità di pubblicazione dei dati aperti del PNRR, in modo che le attività di [monitoraggio civico](#) siano rese possibili.

Nelle prossime settimane saranno pubblicati altri dossier incentrati sulle priorità trasversali riconosciute nel PNRR: i dati per misurare l'impatto sulle questioni di genere, quelle generazionali, e quelle territoriali.

#DatiBeneComune chiede dunque che si passi finalmente dalla teoria alla pratica, dalle parole ai fatti, dalle buone intenzioni alla vera trasparenza e partecipazione. E con spirito costruttivo, collaborativo, partecipativo, suggerisce come farlo.



## PERCHÉ VOGLIAMO I DATI

Nei primi giorni di febbraio 2022 Il Consiglio dei Ministri ha effettuato una ricognizione della situazione relativa ai principali obiettivi del Piano nazionale di ripresa e resilienza (PNRR) del primo semestre dell'anno evidenziando inoltre che, al 31 gennaio 2022, le Amministrazioni titolari di interventi hanno emanato 113 bandi e avvisi per un importo complessivo pari a circa 27,86 miliardi di euro. Ad oggi risultano aperti 48 bandi per un ammontare di risorse da assegnare pari a 23,17 miliardi.

Nel 2022 l'Italia deve conseguire complessivamente 100 obiettivi per il PNRR di cui 83 milestone e 17 target. Di questi 45 sono da conseguire entro il 30 giugno 2022, a cui è collegata una rata di rimborso di € 24,13 miliardi, e 55 entro il 31 dicembre 2022, per la quale è associata una rata di rimborso pari a € 21,83 miliardi.

In questo contesto mettiamo in dubbio il livello di *compliance* che lo Stato ha nei confronti delle scadenze concordate con Bruxelles, quanto evidenziare come le informazioni che dovrebbero far capire il senso complessivo del PNRR alla cittadinanza siano ancora troppo frammentarie. Questo rischia di essere un problema quando, nei prossimi mesi, partiranno centinaia di cantieri e andrà monitorato lo stato di avanzamento dei lavori. Le comunità si chiederanno ad esempio perché è stato deciso di costruire un plesso scolastico in quel quartiere e con quelle modalità e perché non sono state consultate prima di cambiare la viabilità e la vivibilità di un territorio.

Insomma, se da una parte l'emergenza sanitaria ha dimostrato che è fondamentale comunicare bene le scelte politiche che hanno un impatto diretto e immediato sui cittadini e le cittadine, dall'altra è sempre più importante rendere tutte le informazioni e i dati di pubblico interesse accessibili al mondo della ricerca, del giornalismo, delle associazioni e del grande pubblico.

Ciò che vogliamo evidenziare è che **i dati aperti costituiscono uno straordinario strumento di dialogo** tra istituzioni, amministrazioni e cittadinanza, e permettono a quest'ultima di contribuire con saperi e preferenze non solo al controllo e monitoraggio nell'uso delle risorse pubbliche, ma anche al disegno dell'attuazione.

Questo aspetto non viene evidenziato solo dalla società civile, ma la stessa Commissione Europea, nella [Proposta al Consiglio europeo di](#)



[approvazione](#) del PNRR italiano ha rivolto una chiara raccomandazione al Governo italiano: *“Per garantire la responsabilizzazione dei soggetti interessati, è fondamentale coinvolgere tutte le autorità locali e tutti i portatori di interessi, tra cui le parti sociali, durante l'intera esecuzione degli investimenti e delle riforme inclusi nel piano.”*

Come ha raccomandato l'Osservatorio Civico PNRR in sede di conversione del [decreto-legge 31 maggio 2021, n. 77](#), che disciplina la governance del PNRR e le prime misure di rafforzamento delle strutture amministrative e di accelerazione e snellimento delle procedure alla Camera dei Deputati, la pubblicazione dei dati in formato aperto è una preconditione imprescindibile per favorire il monitoraggio civico del Piano e dei suoi progetti, anche al fine di segnalare distorsioni e proporre eventuali azioni correttive.

**Solo grazie ai dati è possibile attuare un efficace monitoraggio civico**, ovvero l'insieme di attività di accompagnamento e controllo diffuso attivate da associazioni e gruppi di cittadini.

Dati aperti e monitoraggio civico possono rafforzare il legame tra amministrazione e cittadinanza, aiutando la prima ad essere più aperta ed inclusiva, e la seconda a partecipare responsabilmente alla costruzione del futuro di tutte e tutti.

La portata complessiva del PNRR italiano raggiunge un importo complessivo di € 191,5 miliardi da impiegare nel periodo 2021-2026 attraverso l'attuazione del PNRR (di cui 51,4 mld per Progetti in Essere). Risorse che stiamo già spendendo e che dovremo spendere con una straordinaria rapidità. Questo si scontra inevitabilmente con un “sistema Paese” che ha una scarsa capacità di prevenire e contrastare la corruzione negli appalti, un sistema di integrità debole e lacunoso e una scarsità di dati aperti disponibili al pubblico.

L'ultimo Indice di Percezione della Corruzione (CPI) pubblicato a gennaio da Transparency International fotografa una **situazione che, seppur in costante miglioramento da una decina d'anni, è ancora lontana dalla sufficienza**. L'Italia, con una valutazione di 56 punti su 100 si attesta nella parte basse della classifica europea, più vicina a paesi come Bulgaria, Ungheria e Romania, che chiudono la classifica, che ai paesi nordici che invece la dominano.





Nonostante la legge anticorruzione approvata nel 2012 abbia dato una scossa all'Italia, non si può ancora affermare che il problema della prevenzione e del contrasto alla corruzione sia stato risolto, anzi.

Gli appetiti criminali d'altronde non si sono mai placati, come dimostrano in maniera cristallina le recenti truffe perpetrate ai danni delle casse statali per il "superbonus ristrutturazioni", che [secondo una stima di Ance](#) rappresentano addirittura poco meno del 20% dell'intero mercato di crediti fiscali generato dalle misure a sostegno dell'edilizia.

La fragilità del sistema di integrità è accentuata da alcune importanti lacune normative, che Governo e Parlamento non sono riusciti - o non hanno voluto - ancora risolvere.

La mancanza di una regolamentazione delle attività di lobbying ad esempio risulta, nell'ottica del PNRR, un vero disastro: le pubbliche amministrazioni si troveranno a dover spendere oltre 200 miliardi di euro in meno di 5 anni, senza alcuna trasparenza sulle modalità con cui chi decide come assegnare i fondi può incontrare chi quegli stessi fondi richiede.

Un vero e proprio telo scuro che nasconde agli occhi della cittadinanza ogni possibilità di conoscere i processi decisionali e di assegnazione di risorse pubbliche.

Come evidenzia la campagna [Lobbying4Change](#) "senza una legge sul lobbying capace di garantire effettiva trasparenza e partecipazione della società civile, c'è il rischio che la politica finisca per favorire solamente chi è più avvantaggiato nei rapporti con il potere e tutti quei soggetti più pronti ad "accogliere" gli investimenti pubblici, escludendo dal piano di "recupero" per il Paese le categorie maggiormente colpite e i portatori di interessi generali che hanno più limitata capacità di farsi ascoltare".

In effetti, grazie alle pressioni della campagna Lobbying4Change una proposta - seppur timida e con diversi difetti - è stata approvata dalla Camera a dicembre 2021, ma ancora si aspetta il via libera del Senato. Visti i tempi molto ristretti del PNRR, sarebbe auspicabile una accelerazione.

Se sul fronte del lobbying siamo dunque in attesa che qualcosa si muova, anche sul fronte contiguo dei conflitti di interesse il Paese -



e le risorse destinate al PNRR - sono in balia di un preoccupante vuoto normativo. Le poche previsioni di legge esistenti al momento per regolamentare i conflitti di interesse sono infatti molto deboli e hanno scarsa efficacia.

Di fatto, la possibilità di scoprire eventuali conflitti di interesse è lasciata in mano a volenterosi giornalisti di inchiesta e attivisti che, a fatica, tentano di fare luce sulla ragnatela di relazioni che avvolge e immobilizza il Paese. Come nel caso del progetto [“Recovery Files”](#) con cui IRPI Media e The Good Lobby stanno tentando di ricostruire i fili che collegano l'erogazione delle risorse del PNRR agli attori dei singoli progetti.

Altra questione rimasta aperta è quella relativa alla **tutela dei whistleblower**, di coloro cioè che segnalano illeciti di cui vengono a conoscenza.

A fronte di una discreta legge approvata nel 2017, l'Italia si trovava a dover trasporre nel proprio ordinamento la nuova e più completa Direttiva Europea sull'argomento. Peccato che la deadline di dicembre 2021 per recepire la Direttiva sia stata “bucata” e che ora ci troviamo in difetto anche su questo versante.

A segnalarlo è lo stesso Presidente dell'Autorità Nazionale Anticorruzione, Giuseppe Busia, che a inizio 2022 ha [richiamato l'attenzione sul mancato recepimento](#), sottolineando che “ come Anac abbiamo contribuito con gli Uffici del Ministero della Giustizia a predisporre un testo, che ritengo fortemente avanzato. Purtroppo è tutto fermo. Non mi risulta che si sia avviato alcun iter per il recepimento”.

Insomma, la totale mancanza di trasparenza sui rapporti tra decisori pubblici e imprese e l'insufficiente protezione per chi segnala casi di corruzione non sono certo un buon viatico per l'investimento dei circa 200 miliardi di euro messi a disposizione dal PNRR.

Anche per questi motivi la società civile italiana ha dimostrato più volte la ferma volontà di giocare il suo ruolo di *watchdog*, di cane da guardia delle istituzioni, e di cercare quantomeno di colmare questi vuoti normativi e di prassi che non consentono un efficace controllo delle risorse pubbliche.

Il problema è che per farlo... ha bisogno di dati.



Anche su questo fronte, purtroppo, le lacune sono molte: i dati messi a disposizione della cittadinanza per monitorare il PNRR sono pochi, pubblicati in ritardo e spesso in formato non machine-readable.

Il [Global Open Data Index](#) pubblicato da Open Knowledge Foundation posiziona l'Italia al 32esimo posto della sua classifica globale, assegnandole una valutazione del 47% rispetto alla quantità e qualità di open data pubblicati. L'indicatore peggiore per l'Italia è proprio quello relativo al "Government Spending" cioè ai dati in grado di mostrare se il denaro pubblico viene utilizzato in modo efficiente ed efficace.

Sull'insufficienza dei dati pubblici relativi al PNRR pubblicati sul portale "Italiadomani.it" è intervenuta anche l'associazione [Openpolis](#) che a fronte di un'analisi approfondita ha fortemente criticato le modalità di pubblicazione, affermando che "tali dati risultano non solo inutilizzabili ma perfino dannosi per chiunque vi faccia riferimento per cercare di comprendere cosa sta accadendo", arrivando ad invitare il Governo "a prendere atto della situazione e a sospendere la sezione open data, almeno fin quando non ci sarà la sicurezza di poter mettere a disposizione di cittadini e analisti dati affidabili".

Tutte queste ragioni, come spiegato in premessa, hanno portato più di 250 organizzazioni ad aderire alla campagna [#DatiBeneComune](#). La società civile italiana non intende stare a guardare passivamente e rischiare di essere una semplice testimone dell'ennesimo spreco di risorse pubbliche.

Questa volta, per la prima volta, pretende di essere protagonista del rinnovamento del paese, di interpretare pienamente il proprio ruolo di cane da guardia e di collettore di istanze e proposte della comunità.

Per riuscirci ha però bisogno di essere messa nelle condizioni di agire in maniera efficace, grazie a dati accurati, tempestivi, aperti, riutilizzabili.

Quei dati che dettaglieremo nei prossimi capitoli.





# I DATI CHE VORREI

## PROGETTI FINANZIATI DAL PNRR

di [Susanna Ferro](#), [Alberto Pampalone Morisoni](#), [Luigi Reggi](#)

Nelle stime relative al Piano Nazionale di Ripresa e Resilienza si calcola che gran parte dei finanziamenti per i progetti (circa il 70%) provverrà da investimenti pubblici.

Appare quindi essenziale che, al fine di garantire l'*accountability* sociale e prevenire fenomeni corruttivi promuovendo un attento monitoraggio della società civile, venga attuata una vera e propria operazione trasparenza del PNRR italiano.

Con trasparenza si vuole intendere, così come definito ex art. 1 del [d.lgs. 33/2013](#) (c.d. decreto trasparenza) come modificato dal [d.lgs. 97/2016](#), quella *"accessibilità totale dei dati e documenti detenuti dalle pubbliche amministrazioni, allo scopo di tutelare i diritti dei cittadini, promuovere la partecipazione degli interessati all'attività amministrativa e favorire forme diffuse di controllo sul perseguimento delle funzioni istituzionali e sull'utilizzo delle risorse pubbliche"*.

A tal proposito – e in via preliminare – si richiede quindi che venga data piena e completa attuazione a tutti quegli strumenti giuridici che, ai sensi della normativa nazionale, già prevedono obblighi di pubblicazione, affinché una reale operazione trasparenza venga effettivamente garantita.

A titolo meramente esemplificativo si fa riferimento agli obblighi di cui al d.lgs. 33/2013, al [d.lgs. 50/2016](#) (codice degli appalti), all'[art. 50](#) comma 1 del CAD "Disponibilità dei dati delle Pubbliche Amministrazioni", alla [L.190/2012](#), nonché alle previsioni di cui alla [delibera ANAC n. 1310/2016](#) recante *Prime linee guida recanti indicazioni sull'attuazione degli obblighi di pubblicità, trasparenza e diffusione di informazioni*.

In particolare, si richiedono dati sui singoli progetti finanziati con il seguente dettaglio.

**Fase di programmazione** (come sono state prese le decisioni di finanziamento?)



- informazioni sui processi di programmazione degli interventi e sulle relative decisioni pubbliche;
- indicazione degli atti pubblici o pubblicazione di minute o analisi alla base delle decisioni prese;
- informazioni su eventuali Programmi di investimento a cui il progetto si riferisce o è collegato (es. programmi collegati da fonti di finanziamento complementari);
- obiettivi di policy collegati al progetto finanziato;
- decisioni sulle modalità di attuazione (es. bando pubblico vs. individuazione diretta);
- indicazione se il progetto inserito nel PNRR è “nuovo” o già avviato in precedenza.

#### **Fase selezione e attivazione dei progetti (cosa e come viene finanziato?)**

- procedure amministrative di “attivazione” (es. delibere, bandi, contratti o altre procedure che danno origine al finanziamento);
- denominazione del progetto selezionato;
- Codice Unico Progetto (CUP);
- descrizione e obiettivi del progetto (descrizione, in un linguaggio comprensibile ai non addetti ai lavori, degli obiettivi e attività previste dal progetto);
- amministrazione titolare/ruolo (attuatore/intermediario);
- enti locali coinvolti/ruolo (soggetto attuatore/destinatario finale);
- altri soggetti coinvolti;
- unità di missione;
- missione;
- componente;
- misura correlata;
- tempistica di completamento;
- fonti finanziarie;
- localizzazione;
- indirizzo e coordinate dei luoghi di realizzazione del progetto;
- indirizzo e coordinate dei soggetti coinvolti;
- indicatori di realizzazione (output) e impatto (outcome).

#### **Fase di attuazione (Come viene realizzato il progetto? Cosa è stato realizzato e cosa rimane da fare?)**

- procedure di gara eventualmente collegate al progetto - Codice Identificativo di Gara (CIG);



- modalità e termini di partecipazione e presentazione delle domande/progetti; ammontare delle risorse disponibili;
- requisiti di accesso dei soggetti proponenti;
- condizioni di ammissibilità dei progetti e delle spese;
- criteri di valutazione;
- forma e intensità dei finanziamenti;
- fasi di attuazione: affidamento, esecuzione, collaudo, rendicontazione spese;
- tempistiche;
- documentazione (atti, contratti, provvedimenti di spesa, etc.);
- codice Fiscale dei soggetti giuridici coinvolti nel progetto;
- avanzamento finanziario (es. pagamenti);
- avanzamento procedurale (es. stato del progetto);
- avanzamento fisico (es. indicatori di realizzazione);
- evidenze sui risultati di eventuali audit amministrativi.

Questi dataset devono inoltre essere resi disponibili in modo da garantire l'[interoperabilità](#) con le principali fonti correlate, di cui si fornisce di seguito un elenco esemplificativo e non esaustivo:

- [Banca Dati Nazionale dei Contratti Pubblici \(ANAC\)](#);
- [Consultazione.gov.it](#);
- [IPA](#);
- MEPA;
- [ParteciPA](#);
- [Registro Imprese](#);
- Registro dei Titolari Effettivi;
- [Registro Nazionale degli Aiuti di Stato](#);
- [OpenDBAP](#);
- [OpenCUP](#).

Infine, si raccomanda di rendere accessibili alla cittadinanza tutti i dataset che le pubbliche amministrazioni acquistano da operatori privati per effettuare il monitoraggio di fenomeni specifici, legati a progetti finanziati dal PNRR, come ad esempio la recente "Acquisizione di dataset per il monitoraggio dei flussi turistici in Italia"<sup>1</sup> da parte del Ministero del Turismo.

---

<sup>1</sup> [Ministero del Turismo](#)



# DATI: COME LI VORREI

di [Andrea Borruso](#), [Giorgia Lodi](#), [Paola Chiara Masuzzo](#), [Davide Taibi](#)

Quando offriamo un bicchiere d'acqua del rubinetto di casa a un nostro ospite, non diciamo "ecco un bicchiere d'acqua potabile". Allo stesso modo, quando si parla di dati, vorremmo che l'aggettivo "open" fosse sottinteso. Così come diamo per scontato che l'acqua di casa sia potabile, vorremmo poter pensare lo stesso per quanto riguarda i dati pubblici.

Ma i dati, proprio come l'acqua potabilizzata, per essere accessibili e aperti necessitano dell'intervento della politica e dell'amministrazione.

La politica deve decidere l'apertura dei dati di pubblico interesse.

L'amministrazione, di concerto con la collettività, deve in prima istanza individuare e scegliere quei dati che, una volta resi pubblici, possano generare valore per la collettività stessa; la pubblicazione e la conseguente diffusione di questi dati devono poi seguire i più elevati standard internazionali.

## I principi FAIR

Un *framework* concettuale di grande importanza è quello dei principi FAIR, un insieme di linee guida formulate nel 2014 per ottimizzare la pubblicazione, la riutilizzabilità, e la qualità dei dati di ricerca, ma che si può, anzi si deve, estendere a tutti i dati di interesse pubblico. I principi FAIR si possono riassumere come segue:

- **Findable (Trovabile):** i dati devono essere facilmente rintracciabili dagli esseri umani e dalle macchine. Questa proprietà è garantita dall'uso di [identificatori persistenti](#) (PID) - DOI, Handle, URN - e dai [metadati](#) descrittivi, che devono essere registrati in "cataloghi" o in repository indicizzabili anche dalle macchine.
- **Accessible (Accessibile):** i dati devono essere sempre accessibili, persistenti nel tempo, e rintracciabili in rete, attraverso protocolli standard. Qualora non fosse possibile rendere i dati "open", sistemi di autenticazione possono essere usati, a patto che almeno i metadati siano sempre disponibili.



- **Interoperable (Interoperabile):** i dati (e i metadati) devono essere **interoperabili**, ovvero devono poter essere combinati con altri dati e strumenti. Questo vuol dire che il loro formato deve essere [aperto](#) (un *CSV* lo è, un *XLS* no), e che i metadati di contenuto e descrittivi devono essere rappresentati in un linguaggio standardizzato (usare [ontologie](#) e vocabolari controllati, ove possibile).
- **Reusable (Riutilizzabile):** i dati devono essere **riutilizzabili**, per cui deve essere chiaro in che modalità sia dati che metadati possono essere riutilizzati (replicati, usati in contesti diversi, con scopo diverso, etc). Questo significa anche accompagnare i dati con una o più [licenze aperte](#), chiare, accessibili e possibilmente riconosciute a livello internazionale.

I principi FAIR sono, per l'appunto, dei principi. Agnostici da implementazioni tecnologiche, costituiscono delle linee guida per garantire non solo che dati/metadati siano rintracciabili e persistenti nel tempo, ma anche che siano pubblicati tenendo in mente standard internazionali di rappresentazione del dato di alta qualità.

Alcuni di [questi](#), anche referenziati nelle [linee guida nazionali per la valorizzazione del patrimonio informativo pubblico](#), hanno tra i loro obiettivi la valutazione dei requisiti di qualità dei dati nella produzione, acquisizione e loro integrazione, così come il miglioramento della qualità dei dati da esporre in modalità "Open data".

## La qualità dei dati

Anche rispetto ai principi FAIR prima enunciati, si ritiene importante riportare di seguito alcune caratteristiche di qualità dei dati, che vorremmo vedere implementate:

- **Accuratezza:** di solito riferita sia all'accuratezza sintattica sia a quella semantica, consente di verificare se il dato rappresenta correttamente il valore reale del concetto o evento cui si riferisce. Questo facilita il raggiungimento degli obiettivi come definiti nei principi FAIR *Interoperable* e *Reusable*.
- **Attualità:** il dato è del "giusto tempo" (è aggiornato) rispetto al procedimento cui si riferisce. Questa caratteristica di qualità è molto importante, è uno dei grossi problemi che affligge i dati



aperti italiani e può anche riferirsi alla cosiddetta tempestività di aggiornamento del dato. Inutile evidenziare che se i dati non sono attuali, disponibili al giusto tempo, i principi FAIR *Accessible*, *Interoperable* e *Reusable* non riescono a essere pienamente soddisfatti.

- **Completezza:** il dato risulta esaustivo per tutti i suoi valori attesi (ad esempio, un dato nazionale dovrebbe coinvolgere i dati di tutte le sue regioni, o ancora il dato sui contagi nelle scuole dovrebbe contenere lo stato di tutte le scuole individuate per il monitoraggio) e rispetto alle fonti che concorrono alla sua creazione (ad esempio i dati aperti sui bandi pubblici dovrebbero integrare i dati provenienti da tutte le istituzioni che pubblicano bandi). Questa caratteristica di qualità concorre al raggiungimento degli obiettivi *Accessible*, *Interoperable*, *Reusable* e *Findable*, in quest'ultimo caso quando applicata ai metadati. Infatti, anche i metadati, che accompagnano i dati, devono essere il più completi possibile per contribuire a migliorare e facilitare la scoperta dei dati.
- **Disponibilità:** il dato è disponibile e quindi viene consentito l'uso dello stesso da parte di utenti (autorizzati) e dalle applicazioni. Tale caratteristica rappresenta l'essenza di quanto indicato nel principio FAIR *Accessible* ma può essere di particolare rilevanza anche per il principio *Findable*: senza disponibilità dei metadati che descrivono i dati, scoprirli diventa molto più complesso, interpretarli, a volte, impossibile.
- **Portabilità:** anche uno dei diritti sanciti dal [regolamento europeo sulla protezione dei dati personali](#) (GDPR), consente di valutare la capacità del dato di essere trasportato da diversi sistemi in un formato strutturato e leggibile da una macchina. In questo caso, solo se il dato è interoperabile (*Interoperable*), si può effettivamente garantire tale caratteristica.
- **Conformità:** questa caratteristica di qualità consente di valutare se il dato aderisce a standard, convenzioni, regolamenti (ad esempio come possono essere ontologie e/o vocabolari controllati). Questa caratteristica facilita il raggiungimento degli obiettivi previsti per i principi *Interoperable* e *Findable*, in questo ultimo caso se i metadati per esempio seguono standard condivisi anche in contesti non necessariamente nazionali.

Si vuole infine evidenziare la necessità di uniformarsi alle linee guida nazionali per la valorizzazione del patrimonio informativo pubblico che, attraverso le [azioni riportate](#), raccomandano anche [piccoli](#)





[accorgimenti](#) (tra cui, ad esempio, specificare il carattere separatore quando si pubblicano CSV), utili a garantire un processo di gestione dei dati di qualità.

## Riferimenti utili

1. DECRETO LEGISLATIVO 24 gennaio 2006, n. 36 - Attuazione della direttiva (UE) 2019/1024 relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico che ha abrogato la direttiva 2003/98/CE ([link](#))
2. DECRETO LEGISLATIVO 7 marzo 2005, n. 82 - Codice dell'Amministrazione Digitale ([link](#))
3. Defining Open in Open Data, Open Content and Open Knowledge ([link](#))
4. Direttiva (UE) 2019/1024 del Parlamento europeo e del Consiglio, del 20 giugno 2019, relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico ([link](#))
5. FAIR Cookbook ([link](#))
6. ISO/IEC 25012 | SQuaRE: requisiti della qualità di sistemi e valutazione di prodotti ([link](#))
7. Linee guida nazionali per la valorizzazione del patrimonio informativo pubblico ([link](#))



# ETICA DEI DATI

di [\*Valentina Bazzarin\*](#), [\*Diletta Huyskes\*](#), [\*Paola Chiara Masuzzo\*](#)

Fare in modo che i dati e quello che producono siano etici va ben oltre la conformità: è un atto di coraggio, e un po', forse, anche di ribellione.

La [\*data science\*](#) ha evidenti impatti sulla società e sulla vita pubblica. Per questo le sfide legate all'etica devono essere tenute in considerazione nel dibattito e soprattutto nelle politiche e nel governo dei dati e delle infrastrutture digitali e tecnologiche.

L'[\*etica dei dati\*](#), come campo specializzato dell'etica tecnologica, studia e valuta le fasi di **generazione**, **registrazione**, **cura**, **elaborazione**, **condivisione** e **utilizzo** dei dati e delle pratiche corrispondenti, comprese le interazioni tra hardware, software e i dati stessi.

In ambito economico sono stati individuati 5 principi da osservare nelle organizzazioni e che possono essere mutuati dalle comunità o dai soggetti che operano nell'ecosistema dei dati, come decisori, come produttori o come fruitori.

1. **Governance** del dato: è necessario che sia sempre identificabile e che l'uso ne sia consentito.
2. **Trasparenza**: è un esercizio doveroso per le pubbliche amministrazioni e indispensabile quando, per esempio, si redige una *policy* o i dati vengono utilizzati all'interno di algoritmi.
3. **Privacy**: quando trattiamo dati personali o dati particolari (biometrici o sensibili) è necessario tutelare la dignità e i diritti delle persone coinvolte, individuare i responsabili e controllare la conformità di ogni trattamento al GDPR.
4. **Intenzioni / obiettivi**: Le intenzioni contano e la minimizzazione dei dati raccolti sono un requisito, quindi prima di iniziare una raccolta o un trattamento è indispensabile chiedersi quali sono gli obiettivi e individuare quali siano i dati necessari al loro raggiungimento.
5. **Outcomes** (risultati/prodotti): se è vero che le intenzioni contano, l'impatto di ogni attività verrà misurato principalmente sui risultati o sui prodotti del trattamento. Per esempio, è sempre necessario valutare se il risultato di



un'analisi o il grafico che lo descrive non contenga dei pregiudizi oppure non possa essere utilizzato per escludere o discriminare dei gruppi o delle persone.

Questi 5 principi guidano ogni passo di un tipico *data science workflow*. Bisogna, in particolar modo, mettere insieme procedure e azioni collettive affinché tutto quello che viene prodotto mediante l'utilizzo di dati pubblici contribuisca alla lotta contro discriminazioni e ingiustizie e alla costruzione di un tessuto sociale più equo: insomma, che **risolva problemi anziché crearne di nuovi**.

Le fasi di un processo di *data science* che dovrebbero essere sempre guidate dai 5 principi possono essere riassunte come segue:

- La **collezione** di dati, prima, la **selezione** delle sorgenti di dati, dopo, e l'eventuale **integrazione** con fonti esterne, devono essere tutte guidate dalla **trasparenza**; fondamentali a tal scopo sono: spiegazione dei metodi utilizzati, chiara provenienza dei dati tramite metadati descrittivi, comunicazione onesta di limitazioni, tecniche e non.
- L'**estrazione delle informazioni** dai dati, che si può ottenere attraverso:
  - la **selezione** dei dati (per un certo scopo): garantire la diversità e l'equità dei campioni e delle dimensioni selezionati non può mai essere un *afterthought*,
  - l'**aggregazione** dei dati tramite la semplice interrogazione degli stessi: tipiche *query* aggregate tendono a nascondere alcune dimensioni/tendenze (si pensi al [paradosso Simpson](#)) e quindi a fornire informazioni distorte e rafforzare discriminazioni;
  - **analisi** dei dati attraverso modelli statistici, algoritmi di [machine learning](#) (modelli addestrati o meno) e di [deep learning](#); è particolarmente importante valutare l'impatto che tecniche analitiche hanno sulla formulazione degli *outcomes* finali, specialmente quando gli algoritmi sono opachi (si pensi al *deep learning*, o più semplicemente ad analisi effettuate con codice non accessibile, non *open source*) e quando hanno potere decisionale sulla vita delle persone;
  - **visualizzazione** dei dati (prima e/o dopo le analisi); il pensiero etico non riguarda solo le intenzioni ma anche le conseguenze che possono derivare dalla presentazione finale dei dati attraverso dashboard,



grafici, report, storie. Generare valore attraverso una storia informata dai dati è un obiettivo che non può mai prescindere dallo sforzo continuo di creare narrazioni della realtà più giuste, eque, inclusive.

Qualsiasi progetto che richieda, utilizzi o raccolga una grande quantità di dati - specie se nel settore pubblico - dovrebbe preoccuparsi della loro **qualità, coerenza e affidabilità** durante l'intero ciclo di vita del progetto, attraverso un **monitoraggio** continuo e analisi del rischio iterative. A causa della natura mutevole di molti tipi di dati, un simile approccio è indispensabile per riadattare il progetto in base agli impatti osservati. In quest'ottica, è fondamentale che questi progetti siano sensibili al contesto applicativo, prevedendo ad esempio un **coinvolgimento attivo delle comunità e delle persone interessate**.

Adottare un approccio etico all'utilizzo dei dati significa impiegarli per generare impatti positivi: come dicevamo prima, è un approccio che va ben oltre la conformità.



## CREDITS

Questa pubblicazione è stata possibile grazie al supporto di Open Society Foundations e alla collaborazione di ActionAid Italia e Transparency International Italia.

Hanno contribuito alla stesura del testo, offrendo competenza, conoscenza ed entusiasmo

### [Valentina Bazzarin](#)

Esperta in Etica dei Dati. Ha conseguito un dottorato in Psicologia Generale e Clinica all'Università di Bologna e attualmente insegna Psicologia Cognitiva nel programma USAC. E' formatrice in corsi professionali di Business Intelligence, Open Data, Open Science e Comunicazione Strategica e basata sui dati. Lavora inoltre come consulente ed esperta esterna per vari soggetti pubblici e privati, in Italia e all'estero.

### [Andrea Borruso](#)

Esperto in Sistemi Informativi Geografici e Open Data. Ha lavorato nella progettazione di GIS e nel data processing di dati spaziali, sia per la Pubblica Amministrazione, che per aziende. Al momento è consulente per il supporto specialistico sui dati aperti. È presidente di onData un'associazione non-profit che promuove l'apertura dei dati pubblici per renderli accessibili a tutti.

### [Davide Del Monte](#)

Ha lavorato oltre 10 anni per Transparency International Italia, di cui è stato direttore esecutivo dal 2015 al 2020. Dal 2021 collabora con diverse organizzazioni no profit che si occupano di tematiche sociali, di trasparenza e di diritti umani. Nel 2019 ha fondato l'associazione info.nodes di cui è attualmente presidente. Da marzo 2021 collabora con onData come coordinatore della campagna #DatiBeneComune.

### [Susanna Ferro](#)

Lavora per Transparency International Italia dal 2013 per cui si occupa delle attività di advocacy dell'associazione. Gestisce e coordina diverse campagne e iniziative volte a migliorare il livello di



integrità in politica e la trasparenza e accessibilità delle informazioni pubbliche per incrementare i livelli di accountability delle istituzioni. Si occupa inoltre di ricerca e data analysis sul finanziamento alla politica e i conflitti di interesse.

### [Diletta Huyskes](#)

Laureata in Filosofia e dottoranda in Sociologia Digitale all'Università di Milano. È la responsabile Advocacy di Privacy Network, associazione nata per la tutela dei diritti digitali con cui ha recentemente progettato l'Osservatorio Amministrazione Automatizzata. Ha lavorato presso la Fondazione Bruno Kessler come assistente ricercatrice sul tema dell'etica e la protezione dei dati in alcuni processi di trasformazione digitale. Studia l'intelligenza artificiale e le nuove tecnologie da una prospettiva etica e politica, concentrandosi sulle disuguaglianze soprattutto di genere e l'uso di algoritmi e software per automatizzare i processi decisionali nella sfera pubblica.

### [Giorgia Lodi](#)

Ha conseguito il dottorato di ricerca in Informatica presso l'Università di Bologna nel 2006. Attualmente è tecnologa presso l'Istituto di Scienze e Tecnologie della Cognizione (ISTC) del Consiglio Nazionale delle Ricerche (CNR) - Laboratorio di tecnologie semantiche (STLab). In questo contesto, è referente privacy dell'Istituto e coordina/partecipa a diversi progetti nazionali ed europei con un focus su apertura di dati secondo il paradigma dei Linked Open Data e qualità dei dati.

In passato ha svolto attività di consulenza per l'Agenzia per l'Italia Digitale (AgID), dove ha lavorato in aree quali open government data, Linked Open Data, Semantic Web, Big Data, sviluppando tra gli altri la rete nazionale di ontologie per la pubblica amministrazione. È stata membro della task force nazionale per l'utilizzo dei dati contro l'emergenza Covid-19 istituita nel 2020 dall'allora Ministro per l'innovazione tecnologica e la digitalizzazione, in accordo con il Ministero della Salute.





### [Paola Chiara Masuzzo](#)

Ha conseguito un Dottorato in Bioinformatica presso UGent (Belgio) nel 2016. Durante gli studi è stata Research Data Alliance fellow e ContentMine fellow. Attualmente è una data scientist a tempo pieno, evangelista ed esperta Open Science, socia onData e ricercatrice indipendente presso IGDORÉ.

### [Alberto Pampalone Morisani](#)

Expert accountability&transparency per Action Aid Italia. Ha lavorato per la cooperazione internazionale, sia in Italia che all'estero (Perù, Sud America), per poi collaborare come legal&policy officer con le pubbliche istituzioni, dapprima presso il Dip. politiche europee della Presidenza del Consiglio e, in seguito, per l'EASO (Ufficio della Commissione europea di supporto all'asilo e alla protezione internazionale) presso il Ministero dell'Interno, dove si è occupato, tra gli altri, di trasparenza, FOIA, accesso agli atti e accessi civici generalizzati.

### [Luigi Reggi](#)

Ha esperienza di ricerca e analisi delle politiche pubbliche in tema di trasparenza, digital government e open data. Da oltre 15 anni si occupa di fondi europei presso i centri di competenza nazionale. PhD in Public Administration and Policy presso la State University of New York at Albany (USA).

### [Davide Taibi](#)

Da più di 15 anni ricercatore presso l'Istituto per le Tecnologie Didattiche del Consiglio Nazionale delle Ricerche, e dal 2016 docente per l'insegnamento di Tecniche per la gestione degli Open Data, nel corso di studi in Informatica dell'Università degli Studi di Palermo. I suoi principali interessi di ricerca riguardano: mobile learning, augmented reality for education, Learning Analytics, Semantic Web e Linked Data per l'e-learning e gli standard per la progettazione dei processi educativi. Dal 2019 coordina due progetti finanziati dalla comunità europea sul tema della Data Literacy.



# GLOSSARIO

## Dati di tipo aperto

I dati che presentano le seguenti caratteristiche: 1) sono disponibili secondo i termini di una licenza o di una previsione normativa che ne permetta l'utilizzo da parte di chiunque, anche per finalità commerciali, in formato disaggregato; 2) sono accessibili attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, in formati aperti ai sensi della lettera l-bis), sono adatti all'utilizzo automatico da parte di programmi per elaboratori e sono provvisti dei relativi metadati; 3) sono resi disponibili gratuitamente attraverso le tecnologie dell'informazione e della comunicazione, ivi comprese le reti telematiche pubbliche e private, oppure sono resi disponibili ai costi marginali sostenuti per la loro riproduzione e divulgazione.

Fonte: [Codice dell'Amministrazione Digitale](#)

## Data science

La scienza dei dati è l'insieme di principi metodologici (basati sul metodo scientifico) e di tecniche multidisciplinari volto a interpretare ed estrarre conoscenza dai dati attraverso la relativa fase di analisi da parte di un esperto (Data Scientist).

Fonte: [Wikipedia](#)

## Dataset

Una qualsiasi collezione di dati organizzata. 'Dataset' è un termine flessibile e può fare riferimento o ad una intera banca dati, o ad un foglio di calcolo o ad altri dati su file, oppure ad un insieme correlato di risorse di dati.

Fonte: [Open Data Handbook](#)

## Deep learning

Si intende un insieme di tecniche basate su reti neurali artificiali organizzate in diversi strati, dove ogni strato calcola i valori per quello successivo affinché l'informazione venga elaborata in maniera sempre più completa.



Fonte: [Wikipedia](#)

## Formato aperto

Un formato di dati reso pubblico, documentato esaurientemente e neutro rispetto agli strumenti tecnologici necessari per la fruizione dei dati stessi.

Fonte: [Codice dell'Amministrazione Digitale](#)

## Identificatori persistenti

Un identificativo persistente (PI o PID) è un riferimento unico e duraturo nel tempo per un documento, un file, una pagina web o altro oggetto digitale.

Fonte: [Università Cattolica del Sacro Cuore](#)

## Interoperabilità

In ambito informatico, la capacità di sistemi differenti e autonomi di cooperare e di scambiare informazioni in maniera automatica, sulla base di regole comunemente condivise.

Fonte: [Piano Triennale per l'informatica nella Pubblica Amministrazione](#)

## Licenza

Uno strumento legale attraverso cui chi detiene i diritti sui dati, può definirne l'uso che se ne può fare. Una licenza *open*/aperta per i dati è quella che consente a chiunque di accedervi, utilizzarli, modificarli e condividerli a qualsiasi scopo, anche commerciale, avendo al massimo come limite misure che ne preservino la provenienza e l'apertura.

Fonte: [Open Definition](#)

## Machine learning

L'apprendimento automatico (\*machine learning\*) è una variante alla programmazione tradizionale nella quale in una macchina si predispone l'abilità di apprendere qualcosa dai dati in maniera autonoma, senza istruzioni esplicite.



Fonte: [Wikipedia](#)

### *Machine Readable*

Un documento è leggibile meccanicamente se ha un formato di file strutturato in modo tale che le applicazioni *software* possano agevolmente identificarlo, riconoscerlo ed estrarne dati specifici.

Fonte: [Direttiva \(UE\) 2019/1024 relativa all'apertura dei dati](#)

### Metadati

Informazioni riguardo un [dataset](#) come il suo nome e descrizione, metodi di raccolta, autore, area geografica e periodo storico a cui si riferisce, licenza, data e frequenza di rilascio, ecc..

Fonte: [Open Data Handbook](#)

### Monitoraggio civico

Il monitoraggio civico è una forma di cittadinanza attiva in cui gruppi di persone o singoli cittadini condividono metodi, strumenti e attività allo scopo di verificare come viene speso il denaro pubblico. Permette inoltre di migliorare il livello di trasparenza delle procedure pubbliche, attivando forme efficaci di controllo sociale e di confronto collaborativo con le istituzioni. In questa pagina trovate le attività di monitoraggio che facciamo anche in collaborazione con i cittadini.

Fonte: [Transparency International Italia](#)

### Ontologia

Nella progettazione di software o di database fondati su programmi che gestiscono basi di conoscenza, un'ontologia informatica viene stabilita strutturando gerarchicamente un insieme di concetti e definendone e descrivendone le relazioni semantiche.

Fonte: [Lessico del XXI Secolo | Treccani](#)

### PNRR (Piano Nazionale di Ripresa e Resilienza)

Il Piano Nazionale di Ripresa e Resilienza (PNRR) prevede un pacchetto di investimenti e riforme articolato in sei missioni. Il Piano



promuove un'ambiziosa agenda di riforme, e in particolare, le quattro principali riguardano:

- pubblica amministrazione;
- giustizia;
- semplificazione;
- competitività.

Fonte: [Ministero Sviluppo Economico](#)