Digeat N.7 - 19 Settembre 2025

Il labirinto della trasparenza: open data, quality assurance e diritti esclusivi

Di Adriana Augenti

Abstract

L’era digitale ha fatto degli open data uno strumento chiave di trasparenza, ma la qualità costa: validazione, pulizia e metadati trasformano dati grezzi in risorse FAIR, creando valore e, spesso, presupposti giuridici per diritti esclusivi. Nasce così il “labirinto della trasparenza”: più rendiamo i dati utili, più rischiamo di chiuderli nei processi che li producono. Questa riflessione si propone di superare la logica binaria aperto/chiuso con una trasparenza stratificata che calibra i livelli di accesso sul valore aggiunto. L’obiettivo non è “uscire” dal labirinto, ma imparare a navigarlo con strumenti di governance più maturi, riconoscendo la trasparenza come valore negoziabile e dunque sostenibile.

Indice

Se una notte d’inverno un cittadino, cercando trasparenza nei dati pubblici, scoprisse che ogni numero nasconde il suo algoritmo, ogni tabella la sua metodologia segreta, ogni grafico il suo processo proprietario, avrebbe trovato il dato che cercava, ma probabilmente avrebbe perduto la strada che vi conduce.

Benvenuti nel labirinto della trasparenza.

L’equazione impossibile: apertura vs. qualità

L’era digitale ha consacrato la trasparenza come principio di governance: **il valore di un dato sta nella qualità**. Come nella caverna di Platone, scambiamo spesso i dati per fatti neutri, ma ogni numero è l’ombra di una scelta, ogni tabella il riflesso di un’interpretazione.

> Nel labirinto, ogni passo verso la trasparenza apre nuove zone d’ombra.

Il mito del dato neutro

L’approccio iniziale al movimento *open data* è stato caratterizzato da un positivismo implicito, che considera i dati come “fatti” oggettivi in attesa di essere semplicemente rilasciati. Questa concezione del “dato grezzo” (*raw data*) come entità neutra e preesistente è, tuttavia, un mito epistemologico che oscura la natura intrinsecamente costruita di qualsiasi informazione.

La filosofia della scienza ha dimostrato la *theory-ladenness* dell’osservazione[1]: ciò che viene misurato dipende dalle ipotesi e strategie dell’osservatore. I dati non “parlano da soli”, ma richiedono interpretazione per acquisire significato. La loro pubblicazione grezza non realizza la trasparenza, ma può creare un data swamp inutilizzabile.

> L’assenza di neutralità del dato giustifica la necessità della *quality assurance* come attività di creazione di valore essenziale.

La *quality assurance* come processo creativo

Il processo che trasforma il dato grezzo in risorsa utilizzabile – la *data curation* – è un’attività intellettuale e creativa che aggiunge valore sostanziale a un dataset.

Le attività di *curation* includono la selezione, l’organizzazione, la pulizia, la validazione, l’arricchimento con metadati e la contestualizzazione, trasformando i dati grezzi in “informazione azionabile” (*actionable intelligence*). **Questo processo mira a rendere i dati conformi ai principi FAIR (*Findable, Accessible, Interoperable, and Reusable*)**, che sono diventati **lo standard di riferimento per la gestione dei dati di ricerca e degli *open data* di alta qualità**.

Raggiungere la conformità FAIR richiede sforzo intellettuale: creazione di metadati, standardizzazione, documentazione e “de-gergalizzazione” implicano scelte discrezionali e competenze specialistiche.

> Questo lavoro di *curation* ha anche una dimensione etica fondamentale: dati non curati possono perpetuare discriminazioni attraverso linguaggio non inclusivo, categorie binarie che escludono identità non conformi, terminologie che riflettono bias di genere storici, o risultare inaccessibili a persone con disabilità o minori competenze digitali.

L’adozione di un linguaggio inclusivo, formati accessibili e standard di usabilità universale rappresenta un ulteriore investimento intellettuale nella qualità del dato, sostenuto dalle Linee Guida AgID, dall’European Accessibility Act (Direttiva (UE) 2019/882) e dal principio di non discriminazione dell’art. 21 della Carta dei Diritti Fondamentali UE.

È proprio in questa aggiunta di valore che si innesta il problema giuridico. Le attività di *curation* corrispondono quasi perfettamente ai presupposti per la tutela della proprietà intellettuale previsti dalla normativa europea sulle banche di dati. La Direttiva 96/9/CE prevede una duplice protezione:

1. Una tutela tramite diritto d’autore per le banche di dati che, “per la scelta o la disposizione del materiale, costituiscono una creazione dell’ingegno propria del loro autore”.
2. Un diritto *sui generis*, indipendente dal *copyright*, per le banche di dati la cui costituzione ha richiesto un “investimento sostanziale, qualitativo o quantitativo, per l’ottenimento, la verifica o la presentazione del loro contenuto”.

> Le attività di *data curation* – come la validazione dei dati, l’arricchimento con metadati, la standardizzazione e la contestualizzazione – possono essere direttamente ricondotte ai concetti legali di “verifica” e “presentazione” del contenuto. Di

> conseguenza, il processo volto a garantire la qualità e l'utilità di un open dataset è lo stesso processo che può generare un diritto di proprietà intellettuale tutelabile.

Ma quando questo principio si applica ai dataset pubblici, emerge una contraddizione fondamentale: l'ente pubblico che investe risorse significative nella qualità del dato può rivendicare diritti esclusivi su informazioni che dovrebbero essere, per loro natura, pubbliche?

La stratificazione del valore

Un dataset non è un'entità monolitica, ma evolve attraverso una serie di fasi che aggiungono progressivamente valore, trasformandolo da una materia prima grezza a un prodotto finito, pronto per l'analisi (*publication-ready*).

Strato 0 – Dato Grezzo: Il punto di partenza, dove i dati vengono raccolti da varie fonti. Spesso non strutturato, incompleto e affetto da errori, inadatto a un'analisi diretta.

Strato 1 – Integrazione: I dati provenienti da fonti diverse vengono combinati e assumono una forma più strutturata.

Strato 2 – Elaborazione: Operazioni di pulizia (*data cleansing*), de-duplicazione, normalizzazione e trasformazione. Questo strato riduce la varianza e il "rumore" nel dataset.

Strato 3 – Arricchimento: I dati puliti vengono arricchiti con metadati, contestualizzati e analizzati per estrarre *insight* significativi. Diventano *actionable intelligence*.

Strato 4 – Presentazione: Le informazioni vengono presentate in formati accessibili e user-friendly, come dashboard o API, pronti per il consumo finale.

Ogni passaggio da uno strato al successivo rappresenta un investimento in termini di lavoro, competenze e tecnologia, e aggiunge un valore misurabile al dataset. Questa visione stratificata evidenzia come la realtà operativa sia più sfumata della contrapposizione binaria tra apertura e chiusura.

> Qui si manifesta il cuore del labirinto: come bilanciare investimento e apertura? La scelta della licenza diventa decisione strategica sui propri asset informativi.

Per il dato grezzo la scelta è semplice: CC0. Per dataset strutturati emerge il dilemma: CC0 o CC-BY per riconoscere la curation? Entrambe le scelte sono giuridicamente legittime, ma riflettono filosofie diverse sulla gestione del patrimonio informativo pubblico[2].

Per un dataset altamente arricchito (Strato 3), frutto di investimenti sostanziali, il dilemma si acuisce. Il quadro normativo presenta una tensione intrinseca tra obblighi di apertura e tutela dell'investimento. L'ente può limitarsi al recupero dei costi marginali previsto dall'art. 7 del D.Lgs. 36/2006, ma il valore creato eccede largamente questi costi. Qui emerge la contraddizione operativa: mentre la Direttiva sulle banche dati riconoscerebbe diritti di proprietà intellettuale per l'investimento sostenuto, la Direttiva Open Data impone la riutilizzabilità. L'art. 11 della L. 633/1941 attribuisce automaticamente questi diritti all'ente pubblico, creando una situazione contraddittoria: l'amministrazione possiede diritti che non può esercitare senza contraddire i propri obblighi di trasparenza.

Il quadro normativo europeo, pur complesso, non risolve questa tensione, ma cerca di affrontarla creando architetture distinte. Da un lato, il Data Governance Act (DGA) crea un framework di *fiducia* per facilitare il riutilizzo di dati del settore pubblico che non possono essere completamente aperti, perché protetti da riservatezza, proprietà intellettuale di terzi o dati personali.

Istituisce la figura di “intermediari di dati neutrali”, che devono fornire il servizio di condivisione in modo sicuro e trasparente, senza possibilità di monetizzare la *curation*. Dall’altro lato, e in modo ancora più pertinente, il Data Act interviene direttamente sullo squilibrio di potere, stabilendo i diritti di accesso e utilizzo dei dati, in particolare quelli generati da prodotti connessi (IoT). Questo regolamento mira esplicitamente a combattere il *lock-in tecnologico*, obbligando i produttori a rendere accessibili i dati agli utenti e facilitando il passaggio tra diversi fornitori di servizi. L’Unione Europea, quindi, tenta di regolare i diversi flussi di dati: quelli aperti, quelli protetti ma condivisibili, e quelli generati privatamente ma il cui accesso deve essere garantito per equità e concorrenza[3].

> La tensione si manifesta nella sua forma più acuta quando processi di elaborazione proprietari generano risultati destinati alla pubblicazione aperta. Un caso emblematico è rappresentato dagli algoritmi di anonimizzazione utilizzati per i dataset sanitari: metodologie spesso brevettate o protette come *trade secrets* vengono applicate a dati che devono essere resi pubblici per finalità di ricerca.

Le tecniche di anonimizzazione – utilizzate anche per i data set pubblici – sono spesso implementate attraverso software proprietari, creando una dipendenza tecnologica: **chi controlla gli strumenti controlla, di fatto, l’accesso ai dati “aperti”**.

La tentazione è di mantenere il controllo sui processi che generano valore, pur proclamando l’apertura dei risultati.

Il costo nascosto della trasparenza

La produzione di dati aperti di qualità è un investimento continuativo. Le stime disponibili **–** prevalentemente statunitensi – quantificano l’impatto della scarsa qualità dei dati con ordini di grandezza significativi per impresa e per sistema Paese, ma **per l’Italia è difficile reperire misurazioni comparabili e aggiornate**. Nel nostro contesto è utile osservare gli effetti concreti nella PA: ritardi, rielaborazioni, difficoltà di interoperabilità, aggiornamenti straordinari e minore riuso. Questo è il “costo nascosto” che le politiche di *curation* vorrebbero ridurre.

Questa contraddizione normativa ha un effetto economico perverso: i fornitori privati di servizi di *data curation*, consapevoli che l’ente pubblico non potrà rivendicare diritti sui dataset risultanti, possono aumentare significativamente i propri prezzi, sapendo che l’amministrazione rimarrà dipendente dai loro servizi proprietari. Il potere contrattuale si sposta così verso i fornitori tecnologici.

La protezione si sposta dai dati ai processi: metodologie di validazione, algoritmi di pulizia e know-how procedurale diventano segreti commerciali tutelati dalla Direttiva UE 2016/943. **Gli enti pubblici possono mantenere conformità formale agli obblighi di trasparenza pubblicando i dataset, preservando però l’opacità sostanziale sui metodi che ne garantiscono la qualità**. Il risultato è una nuova forma di esclusività: non sul dato finale, ma sulla “ricetta” che lo produce.

Questo riconoscimento sta diventando esplicito a livello politico. Il PNRR ha destinato investimenti significativi alla digitalizzazione della Pubblica Amministrazione, riconoscendo esplicitamente che la qualità dei dati pubblici richiede finanziamenti dedicati per infrastrutture, competenze e processi di standardizzazione[4].

> Se le stesse politiche pubbliche che promuovono l’open data riconoscono e finanziano la *curation* come un’attività che richiede investimento economico reale, diventa giuridicamente difficile sostenere che tale spesa non costituisca un “investimento sostanziale” ai sensi della Direttiva europea sulle banche dati.

Il labirinto si chiude su sé stesso: i finanziamenti pubblici per la trasparenza creano le premesse legali per la proprietà intellettuale sui dati pubblici.

Tre vie d’uscita dal labirinto

Il labirinto della trasparenza non si risolve trovando l’uscita, ma imparando a navigarlo. L’approccio binario è inadeguato. **La trasparenza non è uno stato, ma un processo che richiede strumenti di governance sofisticati e sostenibili.**

Tre modelli operativi si stanno affermando nella pratica quotidiana degli enti che gestiscono dati pubblici.

Licenze stratificate: il valore riconosce il valore

La prima soluzione trasforma la licenza in strumento sofisticato: diversi livelli di elaborazione giustificano diversi regimi di accesso. Il dato grezzo (investimento minimo) rimane in pubblico dominio con licenza CC0, quello pulito e strutturato (investimento moderato) adotta una licenza con attribuzione CC-BY, mentre i dataset altamente arricchiti implementano modelli *dual-licensing* o *freemium*.

Un sistema che permette di recuperare i costi dagli utenti commerciali, garantendo accesso gratuito per ricerca e uso civico[5].

Quality commons: quando la comunità cura sé stessa

Una soluzione alternativa distribuisce la responsabilità della qualità in una comunità di stakeholder che collaborano alla curation. Wikidata e OpenStreetMap dimostrano come processi collaborativi possano mantenere standard qualitativi elevati: quando la comunità che beneficia dei dati è anche quella che ne cura la qualità, il dilemma dell’investimento viene aggirato.

Richiede però comunità ampie e competenti, difficilmente replicabile per dataset specialistici pubblici.

Finanziamento pubblico: la qualità come servizio universale

La terza via re-inquadra il problema come questione di finanza pubblica. Se i dati aperti di alta qualità sono un’infrastruttura pubblica essenziale, il loro finanziamento dovrebbe essere a carico della collettività attraverso la fiscalità generale.

> Il PNRR rappresenta già un riconoscimento implicito di questo principio. Questo approccio *public utility* evita completamente il labirinto, trattando la qualità dei dati come investimento strategico nel futuro digitale del paese.
>
> Tuttavia, presenta limiti strutturali evidenti.

Innanzitutto, il PNRR è un investimento una-tantum con scadenza rigida (2026): una volta esauriti i fondi europei, la sostenibilità della qualità dei dati ricade nuovamente sugli enti, che si ritrovano con

infrastrutture costose da mantenere senza risorse dedicate.

In secondo luogo, la logica dell'appalto che governa l'implementazione del PNRR spesso trasforma gli investimenti in trasferimenti di risorse verso consulenti esterni che mantengono proprietà su metodologie, algoritmi e *know-how* sviluppati con fondi pubblici. Gli enti pubblici pagano due volte: per lo sviluppo iniziale e poi per ogni aggiornamento o manutenzione.

Terzo, il focus sulla digitalizzazione delle procedure piuttosto che sulla costruzione di competenze interne crea una dipendenza strutturale dai fornitori tecnologici. La PA acquista soluzioni "chiavi in mano" senza sviluppare le competenze necessarie per gestirle autonomamente, **perpetuando il lock-in tecnologico** piuttosto che contrastandolo come la normativa di settore si prefigge di fare.

Infine, la frammentazione degli interventi – ogni ente, ogni regione, ogni ministero sviluppa le proprie soluzioni – impedisce quella standardizzazione e interoperabilità che dovrebbero essere l'obiettivo primario della digitalizzazione pubblica.

Un esempio concreto è la misura 1.3.1 del PNRR, la Piattaforma Digitale Nazionale Dati, che impone l'esposizione di API per l'interoperabilità dei sistemi pubblici. Nella pratica sta spesso avvenendo che i fornitori sviluppano API tecnicamente conformi ma funzionalmente dipendenti dal mantenimento del contratto: cessato il rapporto commerciale, le API rimangono "esposte e funzionanti" ma non restituiscono più dati, poiché questi sono "strettamente connessi al gestionale" proprietario. L'ente pubblico si ritrova con un'interfaccia vuota e la necessità di ricostruire da zero l'accesso ai propri dati.

La crescente automazione della *data curation* attraverso algoritmi di intelligenza artificiale amplifica questa dipendenza. I sistemi di pulizia, validazione e arricchimento automatizzato dei dati sono sempre più spesso *black box* proprietarie: gli enti pubblici ottengono dataset curati senza comprendere i processi che li hanno generati.

> Quando l'infrastruttura stessa della trasparenza diventa algoritmica e proprietaria, il controllo sulla qualità dell'informazione pubblica si sposta di fatto dalle istituzioni democraticamente responsabili agli attori tecnologici privati.

Il risultato: il PNRR, pur finanziando la trasparenza con risorse pubbliche, spesso genera nuove forme di opacità attraverso la dipendenza tecnologica e la privatizzazione del *know-how* pubblico.

La Trasparenza come valore negoziabile

L'imperativo della trasparenza, per essere efficace, richiede dati di alta qualità, ma la produzione di tali dati genera valore aggiunto tutelabile da diritti di proprietà intellettuale. **L'atto stesso di rendere i dati pubblici utili crea le premesse per la loro potenziale chiusura.**

> Uscire dal labirinto richiede di riconoscere la trasparenza come **valore negoziabile**.
> Ha un costo, e la sostenibilità dipende da governance che bilanci costi e benefici.

La "trasparenza stratificata" calibra l'apertura sul valore aggiunto: dato grezzo aperto, livelli superiori sostenuti da meccanismi che riconoscono costo e valore.

Non si tratta di negare l'ideale della trasparenza, ma di renderlo sostenibile e duraturo. Il labirinto non si risolve trovando l'uscita, ma imparando a navigarlo con strumenti più sofisticati, trasformando la tensione da ostacolo insormontabile a principio di progettazione per una governance dei dati più matura ed efficace.

Il futuro della trasparenza pubblica si costruisce riconoscendo che **l'apertura assoluta può essere nemica dell'apertura duratura**. Solo attraverso modelli che bilanciano ideali e sostenibilità potremo garantire che i dati pubblici rimangano realmente pubblici, non solo 'nominalmente aperti' ma 'funzionalmente inutilizzabili' o 'economicamente insostenibili'.

NOTE

[1] Il concetto di "theory-ladenness of observation" fu introdotto dal filosofo della scienza Norwood Russell Hanson in *Patterns of Discovery* (1958) e successivamente sviluppato da Thomas Kuhn ne *La struttura delle rivoluzioni scientifiche* (1962). La tesi sostiene che non esistono osservazioni "pure" o "neutre": ogni atto di misurazione o registrazione è inevitabilmente influenzato dalle categorie concettuali, dagli strumenti e dalle metodologie impiegate dall'osservatore.

[2] Le licenze per i dati pubblici riflettono scelte strategiche sulla gestione del patrimonio informativo. Creative Commons Zero (CC0) rinuncia a tutti i diritti di proprietà intellettuale garantendo massima interoperabilità, mentre Creative Commons Attribution (CC-BY) richiede solo l'attribuzione dell'autore. AGID 'raccomanda' le CC-BY per tutti i nuovi dati aperti nativi, specificando che l'obbligo di attribuzione segue anche un obiettivo di pubblico interesse: garantire l'affidabilità, la qualità e la tracciabilità del dato pubblico.

[3] La strategia europea sui dati si compone di più pilastri. Tra questi, la Direttiva Open Data (recepita con D.Lgs. 36/2006) impone l'apertura di default. Il Data Governance Act (Reg. UE 2022/868) si occupa della *governance*, creando le procedure per condividere in sicurezza dati protetti tramite intermediari neutrali e promuovendo il "data altruism". Il Data Act (Reg. UE 2023/2854) definisce invece i *diritti di accesso e utilizzo*, affrontando direttamente lo squilibrio di potere e il lock-in tecnologico.

[4] La Missione 1 del PNRR ("Digitalizzazione, innovazione, competitività e cultura") prevede 11,15 miliardi di euro per la digitalizzazione della PA, di cui una parte significativa dedicata al miglioramento della qualità e dell'interoperabilità dei dati pubblici attraverso la Piattaforma Digitale Nazionale Dati e l'implementazione del Polo Strategico Nazionale.

[5] L'Agenzia delle Entrate italiana ha rilasciato i dati cartografici catastali con licenza CC-BY 4.0, trattando il dato geografico come bene pubblico puro sostenuto dalla fiscalità generale. Esempio di segno opposto è quello de L'Ordnance Survey britannico: dati cartografici protetti da Crown Copyright che generano licensing fees per sostenere l'intera operazione. Ogni utilizzo da parte di terzi può espandere i diritti proprietari dell'OS sui nuovi contenuti georeferenziati. Due filosofie opposte della stessa sfida.