

Digeat N.7 - 19 Settembre 2025

“Il dato! Chi era costui?”. Breve viaggio nei dati della pubblica amministrazione

Di Francesco Del Castillo



Abstract

L'articolo propone uno sguardo d'insieme, con alcuni approfondimenti critici, sullo stato e sulla qualità dei dati della pubblica amministrazione italiana, nelle loro declinazioni di dati strumentali alle attività istituzionali, di dati condivisi in ambienti controllati con soggetti autorizzati e di dati condivisi pubblicamente senza alcuna restrizione. L'analisi mira a fare chiarezza e porre le basi per valutazioni più approfondite oltre che a fornire qualche indicazione per distinguere fra storture imputabili ai limiti degli strumenti, alla loro progettazione o implementazione, e storture imputabili all'uso che si fa di quegli strumenti.

Indice

- Introduzione
- I dati condivisi senza restrizione: gli open data
- I dati condivisi in ambienti controllati
- I dati come strumento per le funzioni istituzionali
- Il Catalogo nazionale per l'interoperabilità semantica: un tentativo di standardizzazione
- Conclusioni: mancata percezione dell'importanza del dato?

Introduzione

I grandi *player* internazionali, per i loro scopi imprenditoriali, riescono a “datificare” qualsiasi aspetto della vita dei propri utenti, anche quelli apparentemente distanti dal rapporto di servizio con questi instaurato. Chi abbia esercitato, nei confronti di uno di essi, il diritto di accesso, sancito dall'art. 15 del GDPR, ne ha probabilmente avuto prova, scorrendo la corposa “*copia di dati personali oggetto di trattamento*”. Mettendo da parte valutazioni etiche e considerazioni sulle inevitabili complicazioni di *compliance*, senza dubbio i colossi del mercato globale digitale costituiscono un **riferimento** per l'abilità con cui riescono a **raccogliere ed elaborare dati e tradurli in valore**, che, nel loro caso, si traduce in utile economico che aumenta con l'offerta di servizi personalizzati.

Se tralasciamo la miriade di situazioni intermedie e ci concentriamo sulla situazione nazionale, agli antipodi dei grandi operatori globali, c'è probabilmente **il mondo della pubblica amministrazione italiana**, che persegue **scopi istituzionali essenzialmente slegati da logiche di impresa e di profitto**. Tolti casi esemplari, che pure esistono, a livello diffuso la sensazione è che il settore pubblico viva una sorta di costante **contraddizione** fra le **tecnologie di grido** alle quali aspira e la progressiva **perdita di capacità di raccogliere e memorizzare dati**, con ordine e struttura.

Ci si propone nel seguito di seguire qualche spunto per valutare se questo sentimento abbia qualche oggettiva ragione d'essere, nella consapevolezza che la sfida racchiude in sé il seme del paradosso: come reperire dati a suffragio della tesi della scomparsa dei dati?

A tal fine, giova passare in rassegna i **tipi di dati di cui la pubblica amministrazione è titolare** o con cui ha a che fare. Ci sono dati raccolti ed elaborati **per la conduzione di procedimenti**, attività e funzioni istituzionali in genere e **dati condivisi con l'esterno**. Non sono due categorie disgiunte, anzi, un'amministrazione può condividere dati utilizzati per i propri compiti istituzionali che altre organizzazioni usano a loro volta per i propri. Emerge un'evidente **circolarità dei dati**, raccolti, elaborati, condivisi, rielaborati e ancora condivisi in un ciclo virtualmente continuo, che rende indispensabili, oltre che **dati di qualità**, anche **canali efficaci per reperire e scambiare i dati** di interesse.

I dati condivisi senza restrizione: gli open data

Per *open data*, come definiti nel codice dell'amministrazione digitale (d.lgs 82/2005, art. 1, comma 1, lett. I-bis e I-ter), si intendono quei dati che la pubblica amministrazione rende **disponibili a chiunque**, in **formato aperto** e **senza limitazioni d'uso**^[1]. I dati aperti, solitamente, sono forniti come *dataset*, file (tendenzialmente in formato JSON o CSV) liberamente scaricabili da internet che racchiudono i dati che un'amministrazione è in grado di offrire su un determinato tema.

Il Codice dell'amministrazione digitale prevede (art. 52) che i dati pubblicati dalle amministrazioni si intendano rilasciati come dati aperti e impone (art. 53) alle amministrazioni di pubblicare, nella sezione del sito dedicata alla trasparenza amministrativa, il **catalogo di dati** (anche non aperti), metadati e banche dati di cui sono in possesso con le relative **modalità di accesso**.

Il grande **catalogo nazionale** dei dataset è il portale dati.gov.it che raccoglie, con un meccanismo di federazione di 230 cataloghi, ben 64.368 dataset (alla data del 3 agosto 2025): un numero decisamente ragguardevole.

L'aggiornamento 2025 del [Piano triennale per l'informatica nella pubblica amministrazione](#) include fra gli obiettivi l'**incremento del numero dei dataset pubblicati** nel catalogo nazionale (si veda, ad esempio, la [linea d'azione RA 5.2.3](#)), e pone obiettivi a tutte le amministrazioni, anche locali, di pubblicare un numero significativo di dataset.

Anche se le corpose [Linee guida sugli open data](#), pubblicate nell'agosto 2023^[2], dedicano un'ampia sezione alla qualità del dato, **l'attenzione sembra più orientata alla quantità**. Per esempio, una ricerca del termine "[ordinanze](#)" limitata al tema "governo e settore pubblico" restituisce dataset non facilmente riferibili all'ente che ha emesso le ordinanze, perché figura la regione che ospita il dataset nel suo catalogo, oppure dataset frammentati per anni. Soprattutto, però, emerge con forza la difficoltà di aggregare e rielaborare dataset di dati omogenei messi a disposizione da soggetti pubblici differenti: ogni soggetto, infatti, **espone i dati in modo differente**. Si può osservare, ad esempio, cercando i dataset che riguardano i "[parcheggi](#)", ma non solo.

In termini tecnici, manca uno standard per esporre dati della stessa natura con la stessa struttura (sintattica e semantica). Un problema comune anche agli altri tipi di dati.

I dati condivisi in ambienti controllati

Negli articoli 50 e seguenti del CAD sono descritti, invece, i dati che le pubbliche amministrazioni rendono disponibili, con i dovuti accorgimenti, ad altre amministrazioni *“quando l'utilizzazione del dato sia necessaria per lo svolgimento dei compiti istituzionali”*. Queste, poi, *“procedono all'analisi dei propri dati anche in combinazione con quelli detenuti da altri soggetti”*, per esempio **per fare controlli** e per **dare concretezza a quel principio che recentemente si usa chiamare “once-only”** ma che è costantemente presente nel nostro ordinamento giuridico **da più di mezzo secolo**^[3], prima che si iniziassero a usare nomi anglofoni per i principi base dell'amministrazione.

Fra i dati da condividere “con discrezione”, spiccano le **“basi di dati di interesse nazionale”** (art. 60 del Codice dell'amministrazione digitale) che si contraddistinguono sia per l'autorevolezza dell'**autorità** (tendenzialmente centrale) che le mantiene aggiornate e accessibili, sia per la loro **strategicità** a supporto dell'attività istituzionale dell'intero sistema pubblico. Per non essere troppo teorici, si tratta di anagrafe della popolazione residente, registro delle imprese, archivio dei veicoli e dei patentati, casellario giudiziale ecc.

Lo strumento principale per accedere ai dati condivisi in ambienti controllati è la Piattaforma Digitale Nazionale Dati (PDND – art. 50-ter del Codice dell'amministrazione digitale), che – pur non contenendo alcun dato – rende finalmente possibile a livello tecnico la circolazione di dati fra tutte le articolazioni della pubblica amministrazione, tanto centrale quanto locale, con recente apertura al mondo privato. La PDND espone il catalogo dei servizi di consultazione (o scrittura) disponibili, facilita la conclusione degli accordi di fruizione dei dati e sovrintende, dal punto di vista informatico, all'autorizzazione dei singoli scambi.

Per espressa previsione normativa, tutte le pubbliche amministrazioni sono tenute all'adesione alla PDND, rendendovi disponibili le proprie banche dati e sviluppando i relativi e-service di consultazione. Proprio la pubblicazione di tali e-service è stata oggetto di numerosi avvisi PNRR. In virtù di questo, a inizio agosto 2025, nella [PDND](#) sono presenti **6.074 enti come “erogatori”** e 6.719 come “fruitori” di dati: fra questi, sono attive **18.237 connessioni**. Dei 12.772 e-service pubblicati, 10.258 sono realizzati dai comuni e la quasi totalità di questi non ha più di 2 (due) fruitori attivi^[4]; infine, 382 e-service sono delle pubbliche amministrazioni centrali.

Scorrendo la lista degli e-service più fruiti, tolto il caso particolare di SEND, la piattaforma per le notifiche digitali, che in realtà mette in condivisione funzioni applicative e non meri dati, **emerge come il principio once-only sia ancora lontano dalla sua completa e “automatizzata” realizzazione**: infatti, solo 863 enti sono attivi nella consultazione dell'anagrafe nazionale per l'accertamento della residenza, 788 nella consultazione dell'indice dei domicili digitali INAD. Ancora più significativo è che **meno di 30 enti acquisiscono dall'INPS i dati dell'ISEE**^[5] ed esattamente 5 quelli del DURC^[6].

Anche i dati esposti tramite la PDND sono afflitti dal problema della **mancata standardizzazione delle modalità di presentazione**. Inoltre, gli e-service pubblicati **non sempre hanno documentazione sul significato e il grado di aggiornamento dei dati** condivisi, che, in assenza di standard, sarebbe fondamentale, atteso che si tratta di dati certificativi.

I dati come strumento per le funzioni istituzionali

I dati sono essenziali alla pubblica amministrazione per la conduzione dei propri procedimenti o attività. In ambito locale, per esempio, supportano le funzioni di pianificazione e gestione del territorio o di contribuzione locale, anche nelle attività di

contrasto all'abusivismo e all'evasione. Tali dati sono meno visibili dall'esterno e una loro rassegna risulta meno agevole. Tuttavia, l'esperienza diretta può trovare conferma in qualche dato oggettivo.

Interviene pesantemente sui processi interni, anche se ha risvolti importanti anche per il cittadino, la revisione in digitale dei **pagamenti tramite la piattaforma pagoPA**. Digitalizzare e standardizzare i pagamenti verso la pubblica amministrazione non è il vezzo di un legislatore, ma l'occasione per migliorare, per tutti gli attori coinvolti, **l'intero processo di gestione** del ciclo di vita dei pagamenti. Uno dei benefici meno decantati dalla narrazione che, correttamente, si rivolge alla platea dei cittadini e pone l'accento sui cambiamenti dal punto di vista di chi paga, è quello di **automatizzare le fasi di riconciliazione e rendicontazione** dei pagamenti ricevuti: in altri termini, "chiudere" le posizioni debitorie e imputare i pagamenti ricevuti alle giuste voci di bilancio, nel rispetto delle regole di contabilità pubblica. Senza entrare nei dettagli, la chiave per raggiungere lo scopo è **descrivere in modo completo la posizione debitoria nel momento in cui emerge l'esigenza** di pagamento: il dato contabile è sicuramente noto già in quel momento, va solo memorizzato e incasellato nel giusto posto per abilitare i successivi automatismi. Eppure, il recente [rapporto ANCI sullo stato di digitalizzazione dei comuni italiani](#) mostra (a pagina 31) che, dopo oltre un lustro di uso esclusivo di pagoPA, **la riconciliazione automatica è ancora un lusso per pochi**. Infatti, il 16% dei comuni non la realizza per nessun tipo di pagamento ed è invece più frequente per pagamenti in ambiti specifici quali i tributi o i servizi scolastici, ma non in più di due terzi dei comuni.

Altro esempio di centralità dei dati e della capacità di trattarli è **la sezione "Amministrazione trasparente" del sito** delle pubbliche amministrazioni. Prevista dalla legge 6 novembre 2012, n. 190 sulla repressione della corruzione e dal conseguente decreto legislativo 4 marzo 2013, n. 33 che ne descrive anche i contenuti di dettaglio, nel tempo ulteriormente specificata e ampliata da deliberazioni dell'Autorità nazionale anticorruzione (ANAC), questa dovrebbe essere **una grande vetrina di dati e informazioni** – anche sotto la forma del documento con dati non strutturati – che, in un mondo ideale, potrebbe formarsi da sola, esponendo automaticamente, secondo viste predefinite, una selezione dei dati meticolosamente archiviati negli strumenti di *back office*.

Nell'idea accademica e, forse, nelle intenzioni del legislatore, la richiesta di pubblicare periodicamente (o continuativamente) contenuti specifici dovrebbe **attivare meccanismi di revisione delle modalità di conduzione dei processi** che tengano in considerazione *by-design* l'esigenza di esporre, a un certo punto, alcuni dati significativi del processo stesso. Invece, la trasparenza, per chi naviga la pubblica amministrazione, sembra sempre più **un adempimento a sé stante**, *ex post*, fatto di ricerche spasmodiche di dati non raccolti quando ce n'era l'opportunità.

Anche qui, il paradosso è servito: la possibile leva di efficienza diventa un aggravio procedurale e la richiesta di essere trasparenti in ciò che si fa diventa un adempimento formale autonomo, che consente di superare controlli e risparmiarsi una sanzione, ma che sottrae tempo utile ad altre attività capaci di generare valore.

Il Catalogo nazionale per l'interoperabilità semantica: un tentativo di standardizzazione

Si è detto che nei dataset di *open data* come negli *e-service* della PDND si sente la **manca di standard sintattici e semantici** per lo scambio di dati. Il **Catalogo nazionale per l'interoperabilità semantica** (altrimenti noto come [schema.gov.it](#)), coordinato e realizzato operativamente dall'Istat, tenta di porre rimedio al problema mettendo a disposizione della comunità "ontologie, schemi dati e vocabolari controllati", per fare in modo "*che il formato e il significato dei dati e delle informazioni scambiate siano preservati e compresi in qualsiasi scambio tra le parti*"^[7].

Il senso della standardizzazione semantica è evidente. Se si torna all'esempio degli open data sui parcheggi messi a disposizione da vari enti locali, un operatore economico avrebbe ben interesse a recuperare i dati sull'ubicazione e altre caratteristiche dei parcheggi pubblici, aggregarli e renderli disponibili in una guida turistica o in una mappa interattiva. Per raggiungere lo scopo, avrebbe bisogno che tutti i comuni esponessero le stesse informazioni, con linguaggio uniforme e sempre con la stessa struttura. In effetti, sul Catalogo esiste [un'ontologia per i parcheggi](#), di cui è titolare l'AgID. Da quel che si riesce ad apprezzare, sembra che nessuno dei dataset sui parcheggi presenti sul portale dati.gov.it vi aderisca pedissequamente.

Va sottolineato: il Catalogo per l'interoperabilità semantica **non è per niente una risorsa semplice da consultare** e sembra rivolgersi ad operatori esperti che "masticano" di web semantico, SPARQL, linked open data (LOD), modello di dati RDF ecc. Tuttavia, la lettura del [capitolo quarto della Guida](#) può aiutare anche i meno esperti ad orientarsi e arrivare ai contenuti essenziali delle risorse pubblicate.

Fra le risorse a disposizione, probabilmente, quelle di più immediata fruizione sono i vocabolari controllati, che racchiudono le liste di termini ammessi per descrivere qualche entità. Per esempio, l'INPS mette a disposizione quello dei [documenti di riconoscimento](#) che, questioni di completezza a parte, è sicuramente autoesplicativo [8]. Non sempre è questo il caso: il vocabolario sui [tipi di input e output per i servizi](#) (concetto noto a chi si è occupato di adeguare i siti comunali al nuovo modello nazionale), meriterebbe, al pari di altri, qualche spiegazione sul significato dei termini ammessi, così da guidarne l'uso e sceglierne correttamente uno rispetto a un altro[9].

Conclusioni: mancata percezione dell'importanza del dato?

Quanto fin qui annotato mostra che, in definitiva, l'**apparato normativo** di vario livello a supporto della pubblica amministrazione per raccogliere, gestire e condividere dati di qualità **esiste ed è completo**. Oltre a quanto già citato, lo stesso Codice dell'amministrazione digitale, pur non parlando di qualità, pone con chiarezza ai dirigenti pubblici l'obbligo di raccogliere dati anche quando l'ente dà in appalto (art. 52) o in concessione (art. 50-quater) un servizio, tramite l'inserzione di specifiche clausole contrattuali e pone **accesso e riutilizzo dei dati** fra i parametri di valutazione della performance.

Eppure, in un periodo storico in cui non mancano nemmeno le risorse economiche, **il percorso della pubblica amministrazione verso dati di qualità è ancora lungo**, come lascia trasparire la rassegna innanzi proposta, a tratti impietosamente enfatica su alcune storture, ma con l'intento di aiutare chi volesse approfondire a valutare **se le carenze dipendano da limiti degli strumenti**, della loro progettazione o implementazione, o, piuttosto, **dall'uso che se ne fa**.

A giudizio di chi scrive, le carenze sono da imputare a una diffusa mancata percezione dell'importanza del dato di qualità, sia da parte di chi è chiamato a raccogliere e organizzare i dati sia da parte di chi a questi dovrebbe dare indirizzi. Riguardo questi ultimi, non si può nascondere che puntare primariamente a obiettivi quantitativi ha determinato, sì, la crescita numerica delle voci dei cataloghi, ma anche il proliferare di condivisioni di dati inutilizzate e poco significative. Per raddrizzare il tiro potrebbe bastare poco, magari introdurre nei bandi di finanziamento e nei documenti di indirizzo qualche riflessione esplicita sul "per chi" si condivide e "per quale uso".

Dal punto di vista delle amministrazioni, sempre a detta di chi scrive, sembra prevalere ancora la preferenza per inserire i dati in **documenti discorsivi** senza una controparte di **dati strutturati**. Per cambiare la situazione occorre, operativamente, privilegiare la scrittura di documenti guidata da

software e sistemi nei quali si sono inseriti dati, piuttosto che andarli a recuperare faticosamente, quando servono, nei documenti in cui sono affogati. Nel quotidiano, tuttavia, si osserva che, anche di fronte a un software che mette a disposizione caselle pronte a ospitare dati esatti, l'impulso primitivo è di riempirle in modo approssimativo, giusto per andare avanti, senza riflettere eccessivamente sugli effetti di quelle inserzioni frettolose. Infatti, se queste dovessero confluire, per esempio, nel provvedimento finale, si sarà sempre in tempo ad intervenire per correggere il testo, perché alla fine è quello che, giuridicamente, sembra fare fede. Non si percepisce invece quanti e quali possibilità apra **l'allineamento fra il contenuto dei documenti e le banche dati**, anche solo in termini di **velocizzazione** e **precisione** del lavoro. Al solito, **sensibilizzazione** e **formazione** sembrano l'antidoto migliore a una cattiva abitudine.

Anche dall'alto, però, arrivano segnali confortanti e spinte per un cambiamento di approccio: la recente revisione in chiave digitale ed interoperabile degli sportelli unici delle attività produttive e dell'edilizia è tutta basata sulla convivenza, nei flussi informativi fra gli attori coinvolti, di documenti amministrativi in senso tradizionale e di dati strutturati secondo rigidi schemi standardizzati.

C'è da sperare (e anche impegnarsi per lo scopo) che lo slancio verso la strutturazione del dato non sia annientato dalle altre spinte tecnologiche del momento. I moderni algoritmi, infatti, sono spesso acclamati per la loro possibilità di **scandagliare documenti testuali e ricavarne dati**, anche numerici, strutturati. In questo scenario è concreto il **rischio che le amministrazioni abdichino alla funzione di raccogliere ordinatamente dati** perché "tanto ci pensa la macchina". La macchina magari lasciamola lavorare su documenti vecchi, per creare basi di dati storici, elaborabili e riutilizzabili. Non rinunciamo, invece, a dare ordine e struttura ai dati di oggi, che è sicuro sintomo di **chiarezza di pensiero** e **consapevole governo** dei processi.

NOTE

[1] Il tema degli open data è decisamente vasto e declinato in atti normativi e di indirizzo di respiro internazionale. Per una panoramica si rimanda alla [pagina di contesto del portale nazionale dei dati](#).

[2] Le linee guida sugli open data sono state approvate con [determinazione AgID n. 183 del 03 agosto 2023](#).

[3] Già la legge 4 gennaio 1968, n. 15 indica alcuni fatti, stati o qualità che possono essere autocertificati; l'art. 18 della legge 7 agosto 1990, n. 241, nella sua stesura originale, parla di acquisizione d'ufficio di tutte le informazioni già in possesso della pubblica amministrazione procedente o di altra. Seguono poi le Leggi Bassanini del 1997, il Testo unico della documentazione amministrativa del 2000 ecc. Ovviamente, senza lo strumento tecnico efficace per la circolazione di dati e documenti, tutti i buoni propositi restano lettera morta. Anzi, senza strumento, la semplificazione può diventare aggravio.

[4] Poiché i 10 e-service più fruiti, da soli, realizzano 10.392 connessioni, per forza di cose (e di aritmetica) molte banche dati esposte dai comuni non hanno alcun ente fruitore.

[5] L'ISEE – Indicatore della Situazione Economica Equivalente è alla base della maggior parte delle erogazioni ai cittadini di benefici e sussidi economici, da parte dell'amministrazione centrale e locale.

[6] Il DURC – Documento Unico di Regolarità Contributiva attesta la posizione di un soggetto riguardo i contributi previdenziali e assistenziali.

[7] Dalla "Guida al Catalogo nazionale della semantica dei dati", [paragrafo 1.4](#).

[8] La lista proposta dall'INPS è disponibile anche in [formato CSV](#).

[9] La lista, disponibile anche in [formato CSV](#), sarebbe più significativa se accompagnata dalle definizioni dei termini. Per esempio: quale differenza intercorre fra “certificazione” e “attestazione di identità”? Oppure in che misura “Altra documentazione” è più generico di “Documentazione amministrativa”?