

Interoperabilità e qualità del dato

Di Andrea Piccoli



Rubrica: Gestione documentale 2027

Abstract

Il settimo articolo della Rubrica riprende il tema della interoperabilità dei dati e servizi della pubblica amministrazione, ponendo al centro alcune riflessioni sulla qualità del dato e sulla sua gestione all'interno delle attività e dei procedimenti amministrativi. Prendendo spunto dalle funzionalità della PDND e dalla pubblicazione delle recenti Linee Guida, con Determinazione AgID n. 98 del 19 giugno 2025, si approfondiscono alcuni punti di attenzione nella realizzazione di API per la pubblicazione dei dati e il loro utilizzo nei procedimenti e attività amministrativi, con particolare attenzione alla qualità degli stessi e al loro significato in termini di riutilizzo e interoperabilità.

Indice

- Le linee guida AgID
- Accuratezza
- Completezza
- Aggiornamento
- Coerenza
- Accessibilità
- L'uso della PDND
- Conclusioni

Le linee guida AgID

Le [Linee Guida sull'infrastruttura tecnologica della Piattaforma Digitale Nazionale Dati per l'interoperabilità dei sistemi informativi e delle basi di dati](#), emanate con Determinazione AgID n. 98 del 19 giugno 2025, definiscono all'interno del modello di interoperabilità delle pubbliche amministrazioni quali siano i soggetti che devono e possono usare la piattaforma.

Chiarito e ricordato che la **PDND** non è un raccoglitore di dati, bensì **una piattaforma sicura ed affidabile per la comunicazione e fruizione dei dati di cui le singole amministrazioni restano titolari**, le Linee Guida definiscono i processi di accreditamento, identificazione e autorizzazione assicurati dalla Infrastruttura di interoperabilità PDND, le modalità con cui i soggetti interessati danno seguito alle reciproche transazioni per il tramite dell'Infrastruttura di interoperabilità PDND, nonché le modalità di raccolta e conservazione delle informazioni relative agli accessi e alle transazioni effettuate mediante l'Infrastruttura di interoperabilità PDND.

Da un punto di vista dei principi sulla qualità dei dati, nelle Linee Guida sono introdotti i seguenti punti di attenzione:

– **Accuratezza**: i dati devono essere corretti, privi di errori e coerenti con la realtà che

rappresentano;

- **Completezza:** è necessario garantire che tutti gli attributi rilevanti siano presenti per ogni record;
- **Aggiornamento:** i dati devono essere costantemente aggiornati per mantenere la loro rilevanza;
- **Coerenza:** le informazioni devono essere congruenti tra diverse fonti o dataset.
- **Accessibilità:** devono essere disponibili in formati aperti e leggibili da sistemi informativi interoperabili.

La definizione di dato presente nelle Linee Guida è *“unità minima di informazione digitale, corrispondente a **un valore elementare associato a un attributo di uno schema dati**, che rappresenta un fatto, un evento o una misura in un contesto specifico”*.

L'applicazione di principi sulla qualità del dato, fin dalla sua formazione e gestione, è in carico alla amministrazione titolare, che lo rende fruibile alle altre amministrazioni attraverso l'uso della piattaforma. Il fatto che l'amministrazione titolare si preoccupi dell'applicazione dei principi di qualità del dato è fondamentale per le amministrazioni che li utilizzeranno, perché devono essere sicure della certezza di tali informazioni senza verifiche ulteriori.

Vediamo quindi, nei prossimi paragrafi, come tali principi si traducono in punti di attenzione e modalità pratiche di conduzione.

Accuratezza

L'applicazione del principio di accuratezza dei dati mira a garantire l'assenza di errori e distorsioni negli stessi, in modo da rispecchiare la realtà in modo fedele.

Le prime buone pratiche da seguire sono quelle relative alle tecniche di **verifica dei dati formati e gestiti**, andando ad individuare e attuando le attività di convalida, con controlli formali su formato e range di valori, durante le attività di formazione e gestione del dato. Può essere molto utile, nei contesti in cui è applicabile, la verifica dei dati mediante il riscontro di dataset di riferimento ovvero di altre fonti esterne, impostando degli avvisi automatici in caso di scostamenti significativi (ad esempio su base statistica).

Tra le misure raccomandate c'è quella di **definire, sviluppare e automatizzare, per le diverse tipologie di dati, degli strumenti di monitoraggio degli errori sui dati** nonché di elaborare un **piano di audit specifico per la verifica a campione dell'accuratezza dei dati**.

Completezza

Lo scopo del principio di completezza è quello di assicurare che ogni registrazione di dati contenga tutti gli attributi necessari per soddisfare i requisiti funzionali del caso d'uso.

La prima misura raccomandata è l'**introduzione, all'interno della propria architettura, di servizi di integrazione per un dizionario dei dati**, ovvero di un registro che dettaglia, per ciascun dato, quali sono gli attributi obbligatori e facoltativi. Tale registro può trovare realizzazione anche in una raccolta degli schemi (XSD) utilizzabili anche per la validazione automatica dei dataset pubblicati, comprensiva della verifica dei valori vuoti o mancanti nella pubblicazione.

Una prima buona pratica riguarda la fase di formazione dei dati in cui nei sistemi produttori, o nei flussi di integrazione che vanno a popolare i dati, **devono essere introdotte maschere di acquisizione dati con campi obbligatori evidenziati e passaggi di validazione e completamento**, anche con intervento dell'operatore, nei flussi automatici.

Ovviamente, qualora la mancanza di dati obbligatori divenga significativa (ad esempio se la mancanza di dati obbligatori supera il 5%) e richieda una elevata operatività umana, dovranno essere analizzate le possibili integrazioni aggiuntive anche con sorgenti di dati esterne per integrare gli attributi mancanti.

Aggiornamento

L'applicazione del principio di aggiornamento mira a indirizzare la tempestività nell'aggiornamento dei dati al fine di preservare la rilevanza e l'utilità degli stessi per i soggetti utilizzatori.

La prima misura da attuare è l'aggiunta ai dati pubblicati delle informazioni relative alle date di creazione, di ultimo aggiornamento, di pubblicazione di ciascuna registrazione dei dati. Sulla base di tali informazioni, devono essere definite delle metriche che permettano di monitorare i casi di mancato aggiornamento o di ritardo nell'aggiornamento delle informazioni. È anche una buona pratica definire per ciascun insieme di dati una metrica di obsolescenza delle informazioni e, quindi, gestire in modo opportuno le registrazioni segnalate come obsolescenti sulla base della metrica applicata.

Come nel caso della accuratezza, è buona pratica automatizzare gli strumenti di monitoraggio anche per quanto riguarda l'aggiornamento delle informazioni.

Coerenza

L'applicazione del principio di coerenza ha lo scopo di garantire che informazioni formate, gestite e pubblicate mantengano il loro significato e formato uniforme attraverso le diverse fonti e sistemi.

Le prime evidenti problematiche relative alla coerenza delle informazioni riguardano **i formati di rappresentazione**, si pensi che nella gestione delle date possono essere utilizzati diversi formati (ad esempio, date in dd/MM/YYYY vs. YYYY-MM-DD). È bene, quindi, definire delle regole da applicare in maniera uniforme a tutte le informazioni formate, gestite e pubblicate.

La definizione di ontologie di riferimento per la gestione dei dati pubblicati permette di controllare e assicurare che i dati che rappresentano un determinato fatto trovino la medesima rappresentazione nei diversi dati pubblicati anche per mezzo dell'utilizzo di vocabolari controllati e conseguenti attività di normalizzazione e trasformazione dei dati durante la loro gestione. **I controlli di coerenza delle informazioni possono essere realizzati anche attraverso l'uso di tabelle di riferimento.**

La **definizione di metriche** relative all'incidenza delle riconciliazioni manuali, rispetto alle attività di normalizzazione e mappatura automatizzate, può aiutare a definire un percorso di miglioramento sulla gestione della coerenza dei dati.

Accessibilità

L'applicazione del principio di accessibilità ha lo scopo di rendere i dati pubblicati facilmente reperibili e utilizzabili, sia da persone che da canali di interoperabilità tra i sistemi interni ed esterni.

Oltre a richiamare le Linee Guida sull'accessibilità degli strumenti informatici di AgID, per quanto riguarda le informazioni fruite da persone, alla base delle misure per la realizzazione di questo principio vi è la **scelta di formati aperti e interoperabili per la pubblicazione dei dati e l'uso di OpenAPI[1]**.

Una buona pratica, che favorisce la accessibilità e riutilizzo dei dati, è quella di **rendere pubbliche le informazioni che descrivono i dati pubblicati, i loro schemi e vocabolari di riferimento utilizzati**, rendendo anche disponibili strumenti per la consultazione delle preview dei dati resi disponibili. Anche il tema delle licenze applicate a tali dati pubblicati è un punto importante della comunicazione dei propri dati aperti pubblicati.

L'uso della PDND

L'adozione della PDND consente alle amministrazioni di uniformare i processi di pubblicazione, riducendo i tempi di implementazione degli standard e migliorando la collaborazione tra enti pubblici. Anche dal punto di vista delle misure per l'attuazione dei diversi principi analizzati, la PDND permette di gestire numerosi punti di attenzione.

L'adozione di formati standard (JSON, CSV, XML) per i dati pubblicati, di API RESTfull[2] con licenze aperte (CC-BY[3]), permette di avere buone basi rispetto ai principi di interoperabilità e apertura, accessibilità dei dati e loro riuso.

La PDND, inoltre, offre sia delle funzioni di ricerca sia di consultazione dei dati disponibili e della loro documentazione.

Dal punto di vista della accuratezza e completezza dei dati, la PDND rende disponibili gli schemi dei metadati e le ontologie (secondo lo standard DCAT-AP_IT[4]) e applica una serie di validazioni automatiche integrate nei servizi della piattaforma.

La PDND permette anche una **avanzata gestione sia delle autorizzazioni sia degli accessi associati ai dati pubblici in un vero e proprio registro pubblico delle responsabilità dei dati**; inoltre, offre una dashboard per pianificare gli aggiornamenti dei dati pubblicati, nonché per monitorarne l'utilizzo ovvero per eventuali segnalazioni all'esito dei controlli automatici.

Grazie a un'unica piattaforma centralizzata, si ottiene tracciabilità completa delle modifiche e monitoraggio consolidato delle performance, semplificando attività di controllo, reportistica e manutenzione a lungo termine.

Conclusioni

L'attuazione delle Linee Guida AgID nei progetti di interoperabilità evidenzia molteplici punti di attenzione che hanno il fine di garantire la qualità dei dati interoperabili.

L'adozione della PDND costituisce un'opportunità che permette di **accelerare e condurre in modo attento i progetti di interoperabilità sia in pubblicazione che in utilizzo dei dati pubblici.**

NOTE

[1] [OpenAPI Specification – Version 3.1.0 | Swagger](#)

[2] [What is REST?: REST API Tutorial](#)

[3] [About CC Licenses – Creative Commons](#)

[4] [DCAT-AP_IT v1.1 – Profilo italiano di DCAT-AP | dati.gov.it](#)