



La comunità per i dati aperti pugliesi

DOCUMENTO 1 DI 3

Capire gli open data e i principi FAIR

La cornice: perché e cosa,
prima del come

OPENTUSK

Capire gli open data e i principi FAIR

Documento 1 di 3: la cornice, il perché e il cosa prima del come

A chi serve questo documento

A funzionarie/i, RTD (Responsabili per la Transizione Digitale) e a chiunque, in un ente locale, debba pubblicare dati aperti, anche senza competenze tecniche.

Si rivolgono in particolare agli enti che pubblicano open data seguendo il perimetro nazionale (le [Linee Guida AgID](#)) ed europeo, come comuni, unioni di comuni, università, centri di ricerca e imprese. Per gli enti dell'ecosistema Regione Puglia valgono in aggiunta le regole del Piano di riorganizzazione digitale regionale.

A cosa serve

A capire perché si pubblicano gli open data e cosa rende un dato “buono”, prima di vedere come si carica concretamente sul portale. È la base che dà senso agli altri due documenti:

- **Documento 2, Come costruire un buon CSV** (la preparazione dei dati)
- **Documento 3, Pubblicare su dati.puglia.it** (il caricamento sul portale)

Si legge in circa dieci minuti ed è pensato anche come riferimento da riprendere all'occorrenza.

1. Perché pubblicare dati aperti

Ogni ente pubblico produce dati: elenchi di farmacie, impianti sportivi, contributi erogati, punti di interesse turistico, orari di servizi, etc. Per anni, questi dati sono rimasti chiusi dentro file sui computer degli uffici, utili a chi li aveva creati e a nessun altro.

Aprire i dati significa renderli disponibili a chiunque, in un formato che chiunque possa scaricare e riutilizzare: cittadinanza, giornaliste/i, studenti e studentesse, impresa che ci costruiscono sopra un servizio. Non si tratta soltanto di un adempimento normativo (anche se lo è, in base al Codice dell'Amministrazione Digitale e alle Linee Guida AgID), ma di un modo per restituire valore alla collettività e per evitare che ogni ufficio rifaccia da capo un lavoro che qualcun altro ha già fatto.

C'è però una differenza enorme tra “mettere un file online” e “pubblicare un dato aperto fatto bene”. Un PDF con una tabella dentro è online, ma risulta quasi inutilizzabile, perché nessuno può rielaborarlo facilmente. Un buon dato aperto, invece, è pensato per essere **trovato, compreso e riutilizzato**, anche da un computer e non solo da una persona. È qui che entrano in gioco i principi FAIR.

2. I principi FAIR in parole semplici

I principi FAIR sono nati nel 2016 come linee guida per la gestione dei dati. L'idea di fondo è che un dato, per essere davvero utile, deve avere quattro qualità, riassunte nell'acronimo FAIR:

- **Findable**, cioè trovabile
- **Accessible**, cioè accessibile
- **Interoperable**, cioè interoperabile
- **Reusable**, cioè riutilizzabile

Un dettaglio che spiega molte delle regole che vedremo: i principi FAIR insistono molto sul fatto che i dati siano comprensibili non solo agli esseri umani, ma anche **alle macchine**. Quando cerchi un dataset su un portale, a trovarlo non è una persona che sfoglia, ma un motore di ricerca, un sistema che raccoglie automaticamente i dati da mille fonti. Per questo, conta tanto il modo in cui descriviamo e formattiamo i dati.

F come Findable (Trovabile)

Il primo passo per riusare un dato è riuscire a trovarlo. Un dato trovabile ha tre caratteristiche:

- **Un identificativo univoco e stabile.** Funziona come il codice fiscale per una persona: un'etichetta unica che non cambia nel tempo e che distingue quel dataset da ogni altro.
- **Metadati ricchi.** I metadati sono “i dati che descrivono il dato”: titolo, descrizione, chi l'ha prodotto, quando, di quale zona parla, ogni quanto si aggiorna. Più sono completi, più è facile che qualcuno trovi proprio quel dataset cercando, per esempio, “farmacie provincia di Lecce”.
- **Indicizzazione in un sistema di ricerca.** Il dato va messo in un posto dove i motori lo possono trovare, cioè un portale.

Nel concreto: quando compilerai la scheda di un dataset, i campi come titolo, descrizione, area tematica e identificativo non sono burocrazia, ma sono ciò che rende il tuo dato trovabile.

A come Accessible (Accessibile)

Una volta trovato, il dato deve essere recuperabile in modo chiaro e con strumenti standard, che non costringano nessuno a comprare uno specifico software. Il principio è che il dato pubblico sia raggiungibile da tutte/i con mezzi comuni, come un normale browser web.

C'è poi un punto sottile e importante: i **metadati devono restare accessibili anche se il dato non c'è più**. Se un dataset viene ritirato, la sua scheda descrittiva dovrebbe sopravvivere, così chi lo cercava capisce che esisteva e cosa conteneva, invece di trovare il vuoto.

Nel concreto: è il motivo per cui i portali si “parlano” tra loro. Il portale regionale espone i suoi metadati a quello nazionale, che a sua volta li espone a quello europeo, e così lo stesso dato si trova partendo da più punti.

I come Interoperable (Interoperabile)

I dati raramente vivono da soli: vanno incrociati con altri, caricati in programmi, elaborati. Perché questo sia possibile, devono “parlare una lingua comune”:

- **Formati aperti.** Un formato aperto come il CSV o il JSON può essere letto, elaborato e combinato da qualsiasi software, mentre un formato proprietario lega i dati a un

programma specifico e ne ostacola il riutilizzo. Nel Documento 2 approfondiremo perché il CSV è il “minimo sindacale” di un buon dato aperto.

- **Vocabolari condivisi.** Se per indicare la stessa cosa ognuno usa una parola diversa, i dati non si combinano. Usare categorie e termini standard (per esempio i temi predefiniti del portale, che sono gli stessi del portale nazionale ed europeo, oppure un codice ufficiale ISTAT per indicare un comune) permette di unire dati di fonti diverse.
- **Collegamenti ad altri dati.** Un dato che rimanda ad altri dati correlati vale di più di un dato isolato.

Nel concreto: scegliere il tema corretto da un elenco predefinito, indicare un comune con il suo codice ufficiale invece che con il solo nome, salvare in CSV invece che in Excel: sono tutte scelte di interoperabilità.

R come Reusable (Riutilizzabile)

È l'obiettivo finale: che il dato possa essere riutilizzato in contesti diversi da quello per cui era nato. Perché accada servono tre cose:

- **Una descrizione ricca e accurata** (di nuovo i metadati: più il dato è ben descritto, più è chiaro a cosa serve e quando è affidabile).
- **Una licenza chiara.** Chi trova il dato deve sapere cosa può farci. Una licenza aperta come la Creative Commons CC BY 4.0 – licenza standard prevista dalle Linee Guida AgID - dice: “puoi riutilizzarlo liberamente, anche per fini commerciali, basta che citi la fonte”. Senza una licenza chiara, nel dubbio, nessuno riutilizza il dato per paura di sbagliare.
- **La provenienza (provenance).** Da dove viene il dato, chi l'ha prodotto e come. È ciò che permette a chi lo riusa di fidarsi.

Nel concreto: i campi su editore, autore, licenza e data di aggiornamento servono esattamente a questo. Non sono un di più, ma ciò che trasforma un file in un dato di cui qualcuno si può fidare.

3. Un dato aperto, in pratica, cos'è

Mettendo insieme i quattro principi, un dato aperto fatto bene è:

1. **pubblico e libero da riutilizzare**, senza vincoli e con una licenza chiara;
2. **in un formato aperto** (tipicamente CSV o JSON), apribile e leggibile da chiunque senza software a pagamento;
3. **ben descritto** da metadati completi (chi, cosa, quando, dove, ogni quanto);
4. **identificato in modo univoco e stabile** e pubblicato dove i motori lo possono trovare.

Se manca anche solo uno di questi ingredienti, il dato è meno utile di quanto potrebbe essere. Un Excel pieno di informazioni preziose ma senza descrizione, senza licenza e in formato proprietario è, dal punto di vista FAIR, un'occasione mancata.

Vale la pena ribaltare il messaggio: un dato pubblicato male può essere **peggio di un dato non pubblicato**, perché dà una falsa sicurezza. Chi lo trova crede di avere un'informazione affidabile e ci costruisce sopra decisioni o servizi, e se quel dato è incompleto, ambiguo o

scorretto, l'errore si propaga. La qualità, quindi, non è un dettaglio: è ciò che distingue un dato utile da uno potenzialmente dannoso.

4. La scala della qualità: le “stelle” dei dati aperti

Esiste un modo molto usato per misurare a colpo d'occhio quanto un dato è aperto: la scala a cinque stelle. Più stelle ha un dato, più è riutilizzabile.

Stelle	Cosa significa	Esempio
★	Dato online con licenza aperta, ma in formato chiuso	Una tabella dentro un PDF
★★	Dato strutturato, ma in formato proprietario	Un file Excel (.xls/.xlsx)
★★★	Dato in formato aperto	Un file CSV
★★★★	Come sopra, con identificatori univoci per le singole informazioni	CSV con URI/standard RDF
★★★★★	Dato collegato ad altri dati (Linked Open Data)	Dati interconnessi sul web

La buona notizia è che il **traguardo minimo realistico per un ente è la terza stella, cioè il CSV, condizione minima, tra l'altro prevista dalle Linee Guida AgID**. Non serve arrivare subito alle cinque stelle, serve non fermarsi alle prime due. Un PDF o un Excel da soli non bastano a fare un buon dato aperto, ed è proprio per questo che il prossimo documento è dedicato a come costruire un CSV fatto bene.

Approfondimento ufficiale: i principi FAIR completi, con tutti i sotto-principi, sono consultabili sul sito GO FAIR: go-fair.org/fair-principles

5. Attenzione: non tutto è pubblicabile

C'è un principio di prudenza che viene prima di ogni considerazione tecnica: **aprire un dato non significa che qualunque dato possa essere aperto**. Un dato pubblicato come open data è visibile a tutti, per sempre, e riutilizzabile da chiunque. Per questo non deve contenere informazioni che permettano di identificare singole persone, come nomi e cognomi, codici fiscali, email o indirizzi privati, dati su salute o situazioni personali.

Quando i dati di partenza contengono informazioni di questo tipo (per esempio l'elenco nominativo dei beneficiari di un contributo), prima della pubblicazione vanno rimossi, aggregati o anonimizzati. È una decisione che **non è tecnica, ma di responsabilità dell'ente**: in caso di dubbio si coinvolge il Responsabile della Protezione dei Dati (RPD/DPO) o il referente privacy. La regola pratica è semplice: **nel dubbio, non si pubblica**.

Il modo in cui si tratta concretamente questo aspetto (cosa rimuovere e come aggregare) è spiegato nel **Documento 2**. Qui basta tenere a mente il principio: la pubblicazione di un dato comincia sempre con la domanda “questo dato si può rendere pubblico?”.

6. Pubblicare è un impegno che continua

Un equivoco comune è pensare alla pubblicazione come a un gesto unico: carico il dato e ho finito. In realtà un dato aperto ha un ciclo di vita, e pubblicarlo bene significa prendersene cura anche dopo.

Un elenco di farmacie, di eventi o di contributi cambia nel tempo, tra nuove voci, correzioni e aggiornamenti. Un dato fermo a tre anni fa e mai più toccato perde rapidamente valore e può addirittura ingannare chi lo consulta credendolo attuale. Per questo, tra i metadati che si compilano, c'è anche la **frequenza di aggiornamento**: è una sorta di promessa fatta a chi userà il dato (“questo elenco viene aggiornato ogni anno”, “questo ogni mese”).

Tenerlo a mente fin dall'inizio aiuta a fare scelte sostenibili: meglio pubblicare pochi dataset che si è davvero in grado di mantenere aggiornati, piuttosto che tanti destinati a invecchiare subito. La qualità, ancora una volta, conta più della quantità.

7. Glossario essenziale

Pochi termini ricorrono in tutti e tre i documenti. Eccoli, in breve, come riferimento da riprendere all'occorrenza:

- **Dataset**: un insieme organizzato di dati su un certo argomento (per esempio “le farmacie del comune”). In pratica, di solito, una tabella con le sue informazioni di accompagnamento.
- **Metadati**: i “dati che descrivono il dato”, cioè titolo, descrizione, autore, data, area geografica, frequenza di aggiornamento. Servono a far trovare, capire e fidarsi di un dataset.
- **CSV**: il formato di file aperto più usato per i dati, una tabella salvata come semplice testo, apribile da qualsiasi programma. È il “minimo” di un buon dato aperto (vedi Documento 2).
- **Licenza**: l'indicazione di cosa si può fare con il dato. Per gli open data si usa una licenza aperta (tipicamente [CC BY 4.0](#)), che consente il riutilizzo libero, anche commerciale, citando la fonte.
- **Harvesting**: la “raccolta” automatica con cui un portale di livello superiore (nazionale, europeo) preleva i metadati pubblicati su un portale di livello inferiore (quello regionale). È ciò che fa viaggiare il tuo dato oltre il portale dove lo pubblichi.
- **Codice IPA**: il codice identificativo ufficiale di ogni pubblica amministrazione nell'Indice delle Pubbliche Amministrazioni ([indicepa.gov.it](#)). Serve, tra l'altro, a costruire l'identificativo univoco dei dataset.
- **URI** (Uniform Resource Identifier): un nome univoco che identifica esattamente una risorsa (per esempio un dataset) tra tutte quelle pubblicate. Fa parte di una serie di standard internazionali nati per identificare le risorse, ed è ciò che porta un dato dalla terza alla quarta stella di qualità.
- **RTD**: il Responsabile per la Transizione Digitale, cioè la figura dell'ente che, tra i suoi compiti, segue anche l'apertura dei dati.

8. Come si collega tutto questo al portale regionale

Tutto ciò che hai letto non è teoria astratta: il portale regionale degli open data della Puglia ([dati.puglia.it](#)) è costruito proprio attorno a questi principi. Quando, nel Documento 3, vedrai i campi da compilare, riconoscerai i quattro pilastri:

- l'**identificativo univoco** e i **metadati** del dataset corrispondono al principio *Findable*;
- l'accesso standard e il fatto che i metadati confluiscono sui portali nazionale ed europeo corrispondono ad *Accessible*;
- il **formato CSV** e i **temi e le categorie predefiniti** corrispondono a *Interoperable*;
- la **licenza** e i campi su **editore, autore e aggiornamento** corrispondono a *Reusable*.

In altre parole, ogni regola che il portale ti chiede di rispettare ha una ragione, e quella ragione è FAIR. Sapendolo, compilare quei campi smette di essere un adempimento e diventa una scelta consapevole.

9. Il contesto: Opentusk e il Piano di riorganizzazione digitale

Questi materiali fanno parte dell'iniziativa Opentusk 2026, che promuove la pubblicazione e il riuso degli open data pugliesi. Sul versante regionale, le fondamenta tecnologiche e semantiche sono costruite dal Piano triennale di Riorganizzazione Digitale della Regione Puglia (ultimo aggiornamento con [DGR n. 1646 del 28 novembre 2024](#)). Due dei suoi Obiettivi Realizzativi sono particolarmente legati a questo lavoro.

- **OR_1:** prevede una piattaforma Big Data e un Decision Support System (DSS) che aggrega, anonimizza e fa confluire le informazioni nel patrimonio informativo pubblico, per pubblicarle come open data proprio su dati.puglia.it. È il “motore” che alimenta i dati della Regione.
- **OR_2:** istituisce il Sistema Regionale dei Vocabolari Controllati, per rendere i dati semanticamente interoperabili, attraverso ontologie, nomenclature e vocaboli condivisi, e pubblicabili come Linked Open Data.

Il rapporto con Opentusk è di complementarità. Se gli Obiettivi Realizzativi mettono a terra le fondamenta perché i dati regionali vengano estratti, standardizzati e resi pubblici, Opentusk 2026 fa in modo che questo sforzo non resti fine a sé stesso, stimolando la comunità a riutilizzare quei dati per generare valore, per esempio attraverso le Menzioni Opentusk e la Road to Hackathon. La qualità, la coerenza e l'interoperabilità dei dati sono il ponte tra i due piani.

Per gli enti destinatari di questi documenti: gli Obiettivi Realizzativi sono contesto, non un obbligo. Restano però un riferimento utile, perché adottare vocabolari e standard condivisi, quando disponibili, rende i vostri dati più interoperabili e pronti per il riuso.

10. I prossimi passi

Ora che hai la cornice, i due documenti successivi entrano nel pratico:

- **Documento 2, Come costruire un buon CSV:** come trasformare i tuoi dati (anche se partono da un Excel disordinato) in una tabella pulita, in formato aperto, pronta per la pubblicazione.
- **Documento 3, Pubblicare su dati.puglia.it:** come aderire al portale e caricare concretamente il dataset. Per le schermate di dettaglio rimanda ai manuali ufficiali e alla dimostrazione pratica.

Completano la raccolta due guide operative, da usare quando servono:

- **Guida rapida:** le checklist del CSV e dei metadati e le regole di naming, da tenere accanto mentre si lavora.

- **Come organizzare l'ente per pubblicare:** i ruoli minimi da coprire perché la pubblicazione sia sostenibile e non dipenda da una sola persona.

I principi FAIR sono di GO FAIR (licenza CC BY 4.0). Questo è un documento formativo introduttivo: per gli aspetti tecnici e normativi fanno fede la documentazione ufficiale del portale dati.puglia.it e le Linee Guida AgID.