OPEN TUSK

La comunità per i dati aperti pugliesi

GUIDA OPERATIVA

Prima di iniziare: quale dataset pubblicare

Scegliere e impostare il dataset prima di costruire il CSV

OPENTUSK

Prima di iniziare: quale dataset pubblicare

Scegliere e impostare il dataset prima di costruire il CSV

A chi serve questo documento

A funzionarie/i e RTD di enti locali, in particolare dei Comuni, che devono decidere quali dati aprire e come impostarli, prima di mettere mano al file.

Come gli altri materiali della raccolta, si rivolge 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.

A cosa serve

È la tappa che viene prima del Documento 2. Aiuta a scegliere il dataset giusto, a verificare che valga la pena pubblicarlo e a trovare una struttura di riferimento, così il lavoro tecnico sul CSV parte con il piede giusto.

Fa parte della Cassetta degli Attrezzi Opentusk. Il perché degli open data e i principi FAIR sono nel Documento 1; la costruzione del CSV nel Documento 2; la pubblicazione sul portale nel Documento 3.

1. Perché pubblicare proprio questo dato

Prima ancora di scegliere un formato o una colonna, conviene fermarsi su una domanda semplice: perché questo dato, e a chi serve. Aprire un dato non è un adempimento da spuntare, ma un modo per creare valore, se il dato è utile, di qualità e riutilizzabile.

È il filo che lega tutta la Cassetta: un dato si pubblica perché qualcuno (un cittadino, un'impresa, un altro ufficio) possa trovarlo, capirlo e riusarlo. Sono i principi FAIR del Documento 1, e sono anche il cuore delle regole nazionali, le Linee Guida AgID sull'apertura dei dati, e di quelle regionali, le Linee Guida della Regione Puglia (DGR 217/2024).

Tenere a mente il destinatario aiuta a fare scelte migliori in ogni passo che segue: quali informazioni includere, con quale struttura, con quale frequenza di aggiornamento. Un dato pensato per chi lo userà è quasi sempre un dato fatto bene.

2. Da dove partire: il Paniere regionale

Non serve inventare da zero quali dati aprire. Il percorso Opentusk ha già prodotto un **Paniere regionale**: durante tre laboratori territoriali e una consultazione pubblica, oltre 240 dataset richiesti dal territorio sono stati raccolti e ripartiti per tematica e per competenza (Comuni, Regione, Partecipate e Agenzie). Il quadro completo è nel Report finale della consultazione. Nel 2026 le tappe territoriali Opentusk hanno aggiunto altri 18 dataset, descritti nel Report finale del Paniere regionale 2026: negli allegati sono contrassegnati come nuovi 2026.

Per rendere il Paniere utile a chi pubblica, la Cassetta lo mette a disposizione in tre allegati, divisi per chi ha la competenza sul dato:

- **Allegato 1, dataset di competenza dei Comuni**: l'elenco da cui parte un ente comunale.
- **Allegato 2, dataset di competenza della Regione Puglia**.
- **Allegato 3, dataset di competenza di Partecipate e Agenzie**.

Completano la raccolta l'**Allegato 4** e l'**Allegato 5**, con i dataset che si possono estrarre dai gestionali Maggioli e Halley. Non derivano dal percorso partecipativo: servono a un Comune che usa questi gestionali per individuare i dati che può pubblicare più facilmente, anche in aggiunta a quelli del Paniere.

> **Come usarlo:** parti dai dataset di competenza del tuo ente. Per un Comune è l'Allegato 1, che è di fatto una lista di priorità già indicata dal territorio. Non sei obbligato a pubblicare solo quelli, ma è il punto di partenza più solido, e ti evita di aprire dati che nessuno ha chiesto.

Un principio di buon senso prima di procedere: **si pubblica ciò che è di propria competenza**. Un Comune apre i dati comunali; per i dati di competenza regionale o delle Partecipate se ne occupano gli enti titolari. Aprire un dato che non è tuo genera confusione e duplicazioni.

3. Il dato è davvero disponibile e di qualità?

Scelto il dataset, la seconda verifica è pratica: quel dato ce l'hai, ed è in condizioni utilizzabili? Prima di impegnarti nella preparazione, controlla alcune cose.

- **Disponibilità.** Il dato esiste in una forma da cui puoi partire (un gestionale, un foglio di calcolo, un registro) e non solo sulla carta o nella testa di qualcuno?
- **Qualità.** È aggiornato, completo e attendibile? Un dato vecchio o pieno di buchi va sistemato alla fonte prima di pensare a pubblicarlo.
- **Privacy.** Contiene dati personali (nomi, codici fiscali, indirizzi, informazioni sensibili)? Se sì, vanno rimossi o aggregati prima della pubblicazione. Nel dubbio si coinvolge il referente privacy e non si pubblica. Il tema è ripreso nel Documento 2.

Se il dato non c'è o è troppo disordinato per essere sistemato con ragionevole sforzo, è legittimo rimandarlo e partire da un altro. Vale la regola di fondo della Cassetta: **meglio pochi dataset buoni e mantenuti che tanti pubblicati male e abbandonati**.

4. Esiste già? Evitare le duplicazioni

Prima di costruire un dataset da zero, conviene controllare se un dato uguale o simile è già pubblicato da qualcun altro. Serve a due cose: eviti di rifare un lavoro già fatto, e trovi una struttura di riferimento da riusare. Ecco dove guardare.

- Il **portale regionale**, dati.puglia.it, per vedere cosa è già pubblicato in Puglia.
- Il **portale nazionale**, dati.gov.it, e quello **europeo**, data.europa.eu, che raccolgono i dati di tutte le amministrazioni.
- I due strumenti **AgID**, gratuiti e a supporto delle amministrazioni (rif. 3):

> **Cruscotto Italia**: una piattaforma che raccoglie in un unico posto i principali dati di ogni Comune italiano, presi da fonti istituzionali ufficiali (ISTAT, MEF, Agenzia delle Entrate, ANAC e altre) senza oneri per l'ente. È utile per capire, a colpo d'occhio, quali informazioni su un territorio esistono già ed evitare di ricrearle.
>
> **SIMBA**: una suite pensata per chi produce open data. Permette di cercare nel catalogo nazionale con un assistente conversazionale e, soprattutto, di caricare un file CSV ottenendo un punteggio di qualità e la lista dei problemi da correggere, con l'indicazione di vocabolari e standard da usare. Tornerà utile anche dopo, quando il CSV sarà pronto.

Nel concreto: se stai per creare, ad esempio, una mappa degli incidenti stradali e ne trovi una analoga già pubblicata da un altro Comune e validata a livello nazionale, quella diventa la tua base di partenza, invece di ripartire da zero.

5. Trovare un template di riferimento

Un **template** è semplicemente la struttura di colonne consigliata per un certo tipo di dato: quali campi mettere e come chiamarli. Partire da un template ha due vantaggi: non devi inventare le colonne, e il tuo dato diventa **confrontabile** con quelli di altri enti, che è il senso dell'interoperabilità.

Per i dataset del Paniere, la Cassetta ti dà molto più di un'indicazione. Negli **Allegati 1, 2 e 3** la colonna “Template di riferimento” ti dice a colpo d'occhio, per ogni dato, a che punto sei:

- **Template pronto**: nel documento **Template di riferimento** dello stesso allegato trovi una scheda dedicata con le colonne consigliate e la loro descrizione, e nel pacchetto **Template CSV** dell'allegato un file già pronto da scaricare, aprire e compilare (con anche la riga di intestazione da copiare, per chi preferisce). Se il dato coincide con uno già trattato, la scheda rimanda a quella da usare, anche in un altro allegato.
- **Nessun dataset simile a livello nazionale**: la ricerca è già stata fatta e non esiste un analogo. La struttura la definisci tu, in base ai dati che hai.
- **Non trattato**: la voce del Paniere non descrive un dataset e non ha una scheda.

In ogni caso il template è un punto di partenza da adattare, non una gabbia: aggiungi o togli colonne in base a ciò che il tuo dato realmente contiene.

6. La checklist prima del CSV

Se hai risposto sì a queste domande, sei pronto a passare alla costruzione tecnica del file, descritta nel Documento 2. Se qualche casella resta vuota, conviene chiuderla prima.

- □ Ho verificato che il dato è di **competenza del mio ente** (per un Comune, è nell'Allegato 1)?
- □ Il dato è **realmente disponibile** e in una forma da cui posso partire?
- □ Il dato è **aggiornato e attendibile** a sufficienza?
- □ Ho verificato l'assenza di **dati personali**, o li ho rimossi o aggregati?
- □ Ho controllato che il dato **non sia già pubblicato** (portale regionale, dati.gov.it, data.europa.eu, Cruscotto Italia, SIMBA)?
- □ Ho individuato un **template di riferimento** (allegato di competenza e relativo documento Template di riferimento, con il file CSV) o, in mancanza, un dataset analogo da cui partire?

> **Il percorso completo:** individua il dataset, verifica cosa esiste già, individua un template, prepara e valida il CSV (Documento 2), pubblica sul portale (Documento 3). Questo documento copre i primi tre passi, quelli che si fanno prima di aprire il foglio di lavoro.

7. Riferimenti utili

I richiami tra parentesi quadre nel testo rimandano a questo elenco.

[1] **Linee Guida AgID sull'apertura dei dati:** le regole tecniche nazionali. Linee Guida Open Data AgID

[2] **Linee Guida regionali Open Data (Regione Puglia):** adottate con D.G.R. n. 217 del 4 marzo 2024. delibera e allegato (PDF)

[3] **Cruscotto Italia e SIMBA (AgID):** gli strumenti nazionali per consultare e produrre open data di qualità. Presentazione ufficiale AgID

[4] **Portale nazionale degli open data:** dati.gov.it

[5] **Portale europeo degli open data:** data.europa.eu

[6] **Portale regionale degli open data:** dati.puglia.it

[7] **Paniere regionale e allegati:** il Report finale della consultazione pubblica Opentusk. Report finale (PDF). Per le voci nuove 2026: Report finale del Paniere regionale 2026.

[8] **Allegati e template:** gli Allegati 1, 2 e 3 (dataset di Comuni, Regione, Partecipate e Agenzie, con la colonna Template), i documenti **Template di riferimento** con le schede delle colonne e i pacchetti **Template CSV** con i file pronti da compilare; gli Allegati 4 e 5, con i dataset dei gestionali Maggioli e Halley.

[9] **Gli altri documenti della raccolta:** Documento 1 (open data e principi FAIR), Documento 2 (come costruire un buon CSV), Documento 3 (pubblicare su dati.puglia.it), la Guida rapida e Come organizzare l'ente per pubblicare.

Documento introduttivo e orientativo. Gli strumenti e i riferimenti citati sono di AgID e della Regione Puglia; per i requisiti tecnici fanno fede la documentazione ufficiale del portale dati.puglia.it e le Linee Guida AgID. Gli esempi eventualmente citati hanno scopo puramente dimostrativo.