OPEN TUSK

La comunità per i dati aperti pugliesi

GUIDA OPERATIVA

Guida rapida

Checklist e regole di naming
per la pubblicazione di open data

OPENTUSK

Guida rapida

Checklist e regole di naming, da tenere accanto mentre si lavora

> **A cosa serve questo documento**
>
> È il riassunto operativo dei tre documenti principali, pensato per chi li ha già letti e ha bisogno solo di un promemoria da tenere aperto mentre prepara e pubblica un dataset. Contiene la checklist per il CSV, la checklist per i metadati e le regole per dare un nome ai 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.
>
> *Per il perché di ogni regola, il riferimento resta la documentazione completa: il Documento 1 per i principi, il Documento 2 per la costruzione del CSV e il Documento 3 per la pubblicazione sul portale.*

1. Checklist del CSV

Da spuntare prima di considerare pronto il file dei dati.

Struttura della tabella

- □ Una riga per ogni record, una sola cosa per riga
- □ Una colonna per ogni tipo di informazione, un solo valore per cella
- □ Eventuali raggruppamenti trasformati in colonne (es. `provincia`), non in righe-titolo
- □ Prima riga con intestazioni chiare, minuscole, senza spazi né simboli
- □ Niente celle unite, righe vuote, righe di totale, loghi o titoli

Contenuto

- □ Date tutte nello stesso formato (consigliato `AAAA-MM-GG`)
- □ Numeri senza unità di misura e con separatori coerenti
- □ Valori uniformi dentro ogni colonna (sempre `Lecce`, non anche `LECCE` o `Lecce (LE)`)
- □ Celle vuote dove il dato manca, davvero vuote (nessun valore inventato, nessun `N/D` o `0`)
- □ Valori riportati fedelmente, anche quelli “strani” (semmai segnalati, non corretti)

Formato e privacy

- □ Salvato in **CSV UTF-8**, separatore **virgola**
- □ Accenti corretti (verificati aprendo il file con un editor di testo)
- □ Nessun dato personale (o dati aggregati/anonimizzati previa valutazione del referente privacy)

2. Checklist dei metadati

Da spuntare quando si compila la scheda del dataset sul portale, prima di renderlo pubblico. L'elenco esatto dei campi obbligatori è nella documentazione ufficiale del portale, che prevale in caso di differenze.

- ☐ Titolo chiaro e descrittivo, che inizia con le parole importanti
- ☐ Identificativo con il **codice IPA corretto** come prefisso (la parte dopo i due punti la genera il portale, non va toccata)
- ☐ Descrizione compilata: cosa contiene il dataset, da dove vengono i dati, a cosa servono
- ☐ Area tematica **coerente con il contenuto** (le farmacie vanno in Salute, non in Istruzione e cultura)
- ☐ Licenza aperta indicata (tipicamente **CC BY 4.0**)
- ☐ Editore e organizzazione corretti
- ☐ Data di modifica aggiornata e frequenza di aggiornamento dichiarata (e sostenibile)
- ☐ Riferimento geografico compilato, se il dato riguarda un territorio specifico
- ☐ Dataset tenuto in stato **privato** durante la compilazione, e reso pubblico solo dopo la verifica

> **Ricorda i ruoli:** se sei Curatore, il passaggio da privato a pubblico lo deve fare un Amministratore della tua organizzazione.

3. Regole di naming dei file

Un buon nome di file dice subito cosa contiene, non crea problemi tecnici quando viene scaricato o pubblicato e resta leggibile nel tempo. Le regole sono poche:

- **Tutto minuscolo**, perché maiuscole e minuscole mischiate generano confusione e, su alcuni sistemi, file diversi.
- **Niente spazi**: al loro posto si usa l'underscore _.
- **Niente accenti né caratteri speciali** (à, è, °, €, parentesi, virgole), che possono corrompersi nei passaggi tra sistemi.
- **Contenuto, territorio e anno** nel nome, così il file è riconoscibile anche fuori dal suo contesto.
- **Niente versioni ambigue** come "def", "finale", "ultimo" o "(2)": se serve una versione, meglio una data.

```
✓ farmacie_muro_leccese_2024.csv
✓ eventi_estate_2026_comune_demo.csv
✓ contributi_associazioni_2025.csv
✗ Elenco def (2).csv
✗ DATI FARMACIE aggiornato FINALE.xlsx
```

✗ `però_così_no.csv` (contiene un accento)

Lo stesso criterio vale per i **nomi delle colonne** dentro il file: minuscoli, senza spazi (con l'underscore), senza simboli né unità di misura, come `data_rilevazione` o `importo_euro`.

4. Prima di pubblicare: l'ultimo sguardo

Quando entrambe le checklist sono complete, resta un ultimo controllo d'insieme, che richiede un minuto:

- □ Il CSV si apre correttamente e mostra i dati attesi
- □ La scheda del dataset è completa e senza errori di battitura
- □ Nessun dato personale è sfuggito ai controlli
- □ So chi aggiornerà questo dato, e ogni quanto

Se tutte le caselle sono spuntate, il dataset è pronto per diventare pubblico.

Guida rapida di sintesi: per i dettagli e le motivazioni delle regole si rimanda ai tre documenti principali e alla documentazione ufficiale del portale dati.puglia.it, che prevale in caso di differenze.