

SCHEDA OPERATIVA – SCARICARE, GESTIRE E CONSERVARE I DATI

SEZIONE 1. SCARICO DEI DATI: DA DOVE ARRIVANO E COME SI PREPARANO

1 - INDIVIDUA LA FONTE

Importante distinguere tra:

FONTI INTERNE	FONTI ESTERNE
• Registro elettronico • File excel o google sheets già esistenti • Prove comuni • Questionari già somministrati • Verbali / report interni • Elenchi di partecipazione ai progetti	• Restituzioni invalsi • Open data ministeriali o regionali • Dati comunali o territoriali • Report di enti partner • Questionari o monitoraggi di soggetti esterni

2 - STUDIA IL DATO A CUI TI STAI APPROCCIANDO

- qual è l'**unità di osservazione**: studente, classe, docente, plesso?
- qual è il **periodo di riferimento**: mese, quadrimestre, anno scolastico?
- quali sono le **variabili minime** che ti servono davvero?
- il file contiene **identificativi diretti** inutili per l'analisi?

3 - SCARICA IN UN FORMATO LAVORABILE

- meglio CSV o XLSX
- evitare PDF o DOCX come formato di lavoro
- se il sistema esporta male, creare subito una copia pulita
- non lavorare mai sull'unico file originale

4 - NOMINA I FILE IN MODO RIGOROSO

Esempio: *assenze_studenti_sec1_2025-2026_export_2026-03-15.xlsx*

Esempio regola generale: **contenuto + unità + periodo + data di scarico**

5. CREA CARTELLE DI LAVORO ORGANIZZATE

Esempio:

- *dataset_originali*
- *dataset_puliti*
- *dataset_finali*
- *documentazione*
- *output_grafici_tabelle*

BUONE PRATICHE DI CONSERVAZIONE E CONDIVISIONE

- tenere separati file originali e file di lavoro
- limitare l'accesso ai file con dati identificativi
- usare versioni nominate con data
- evitare invii incontrollati via mail o chat
- condividere, quando possibile, versioni pseudonimizzate o aggregate

Questo è coerente con i principi ribaditi dal Garante: necessità del trattamento per compiti istituzionali, minimizzazione, integrità e riservatezza; per trattamenti statistici o analitici, anonimizzazione o pseudonimizzazione quando possibile.

SEZIONE 2. GESTIONE DEI DATI: COME COSTRUIRE UN DATASET

REGOLE PRATICHE PER COSTRUIRE EFFICACEMENTE UN DATASET

1. USA UN IDENTIFICATIVO STABILE

Idealmente, ogni dataset dovrebbe avere una riga unica e spesso, per identificarla comodamente, si utilizza un identificativo (ID) ad hoc.

Esempi:

- *id_studente*: es. *codice fiscale*
- *id_classe*: es. *2A_2025_Ling*
- *codice_plesso*: es. *codice meccanografico*

Se non hai un ID ufficiale, puoi crearne uno interno pseudonimizzato. Questo è utile anche per ridurre l'uso di identificativi diretti.

Questo, oltre a rendere più facile l'identificazione e il raggruppamento delle variabili, permette anche di far comunicare dataset diversi utilizzando gli identificativi chiave in comune.

2. NON UNIRE NOME E COGNOME COME CHIAVE

È una pessima idea:

- errori di spelling
- omonimie
- problemi privacy

3. UNA COLONNA, UNA SOLA INFORMAZIONE

Evitare:

- *classe_sezione = 2A*
- *esito = promosso con debito*
- *note = assente spesso ma migliorato*

Meglio separare:

- *classe = 2*
- *sezione = A*
- *promosso = 1*
- *debito = 1*
- *le note testuali in un file a parte, se servono davvero*

Inoltre, non conviene usare celle unite, colori o commenti come informazione. Per l'occhio umano sembrano comodi, ma per l'analisi nei fogli di calcoli spesso non sono utili o persino ostacolanti.

4. DECIDI UNA CODIFICA COERENTE DEI DATI MANCANTI

Per esempio:

- cella vuota = dato mancante
- NA = non disponibile
- Valori numerici: meglio non utilizzarli se non supportati esplicitamente dal software come dato mancante, altrimenti creano problemi nei calcoli

La regola fondamentale è decidere una sola codifica supportata dal software utilizzato e documentarla chiaramente.

5. CREA UN FOGLIO "LEGENDA" (METADATI)

Con:

- nome variabile
- significato
- formato
- unità di misura
- codifica
- fonte
- data ultimo aggiornamento
- altre informazioni importanti di contesto legate alle variabili

Questa è una versione semplice di metadati, ma fondamentale per mantenere coerenza e integrare il lavoro in diversi momenti nel tempo.

STRUTTURA IDEALE DI UN DATASET

La regola base è la seguente:

ogni riga = un'unità di osservazione

ogni colonna = una variabile

ESEMPIO 1 – DATASET STUDENTI

Una riga per ogni studente.

Esempio colonne:

id_studente: xxxxx	id_classe: xxx	classe: 2	sezione: A	anno_scolastico: 2025	genere: M	assenze_totali: 8	ritardi_totali: 10	media_italiano: 7.7	media_matematica: 7.2	partecipazione_progetto_x: 1
--------------------	----------------	-----------	------------	-----------------------	-----------	-------------------	--------------------	---------------------	-----------------------	------------------------------

Questo è un esempio di dataset utile per indagare, ad esempio, se esiste una correlazione (NON causazione) tra partecipazione a un progetto e voti di italiano e matematica.

ESEMPIO 2 – DATASET CLASSI

Similmente, un esempio di dataset a livello classe:

id_classe	ordine_scuola	plesso	n_studenti	quota_assenti_frequenti	media_invalsi_italiano	media_invalsi_matematica	n_progetti_attivi
-----------	---------------	--------	------------	-------------------------	------------------------	--------------------------	-------------------

Questo serve quando il livello di analisi non è il singolo studente ma la classe.

È possibile unire i dataset di livello più macro (classi) a quelli più micro (studenti) sfruttando gli identificatori comuni (id_classe, in questo caso), che identificano un solo studente e i dati della sua classe di appartenenza. In questo contesto, si avranno le colonne con i dati delle classi ripetute per ogni studente appartenente a quella classe.

ESEMPIO 3 – DATASET QUESTIONARIO

Una riga per ogni rispondente, una colonna per ogni domanda.

Colonne

id_rispondente	classe	q1_appartenenza	q2_relazione_docenti	q3_stress_scolastico	q4_supporto_compagni	q5_soddisfazione_generale
----------------	--------	-----------------	----------------------	----------------------	----------------------	---------------------------