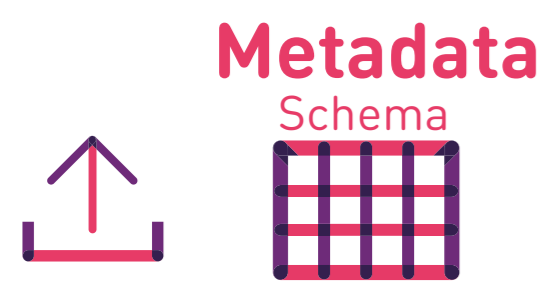
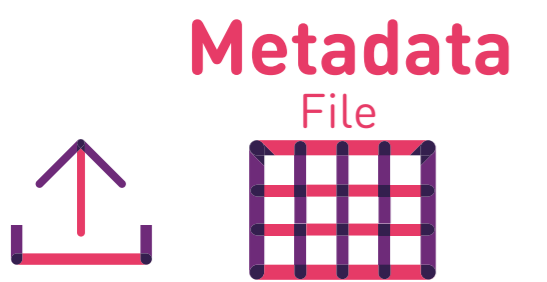


Dataset validation - Metadata



- ✓ • **File format** is Excel or CSV/TSV
- ✓ • Has **all mandatory columns**: StudyStage, ColumnName, Role, Type, Ontology, Unit, UnitOntology, Description
- ✓ • Only **valid types** are given (float, int, string)
- ✓ • At least one line in the schema file in the **role** column contains a value.
- ✓ • **Role** column contains:
 - a) For RAW datasets: one **sampleID** and an optional groupID
 - b) For PROCESSED datasets: **at least one of sampleID, groupID or contrastID**

- ✗ • Unrecognised file format: must be a valid Excel or UTF-8 CSV file.
- ✗ • CSV file must use UTF-8 character encoding.
- ✗ • Schema file must have exactly these columns: list of columns
- ✗ • The following schema columns are invalid: list of columns
- ✗ • The following schema columns are missing: list of columns
- ✗ • Unknown type x found in the Type column: column
- ✗ • Role column is empty
- ✗ • Either sampleID or groupID or contrastID (or any combination) must occur (exactly once), but none of these was present
- ✗ • Role x must occur exactly once
- ✗ • Role x must occur exactly once but it appears y times



- ✓ • **File format** is Excel or CSV/TSV

- ✗ • Unrecognised file format: must be a valid Excel or UTF-8 CSV file.
- ✗ • CSV file must use UTF-8 character encoding.



- ✓ • Metadata file has exactly the **columns specified in the schema file**
- ✓ • For each row in the metadata file, check that the value matches the **type** specified in the schema.
- ✓ • For each row in the metadata file, check that if the schema indicated a **file reference** the field is not empty.
- ✓ • For each row in the metadata file, check that the values are unique if the schema specified a **unique value** in the column.
- ✓ • For each row in the metadata file, check that if the schema indicated a **unit** the field is not empty.
- ✓ • For each row in the metadata file, if the schema specified a unit and the value column is non-empty, check that the **unit** column also contains a value.
- ✓ • For each row in the metadata file, check that if the schema indicated a column as an **ID** the field is not empty.

- ✗ • Metadata file must have exactly these columns: columns
- ✗ • The following metadata columns are invalid: columns
- ✗ • The following metadata columns are missing: columns
- ✗ Invalid data type in the column x at line y: expected r but found 's' of type t
- ✗ • Missing value in the file reference column x at line y
- ✗ • Found duplicate value in the unique column x: y (occurs z times)
- ✗ • Unit is present for a value with unit but the value is missing in column x at line y
- ✗ • Missing unit for a value with unit in column x at line y
- ⚠ • Missing value in the ID column x at line y

Legend



upload



check



validation



possible errors



warnings