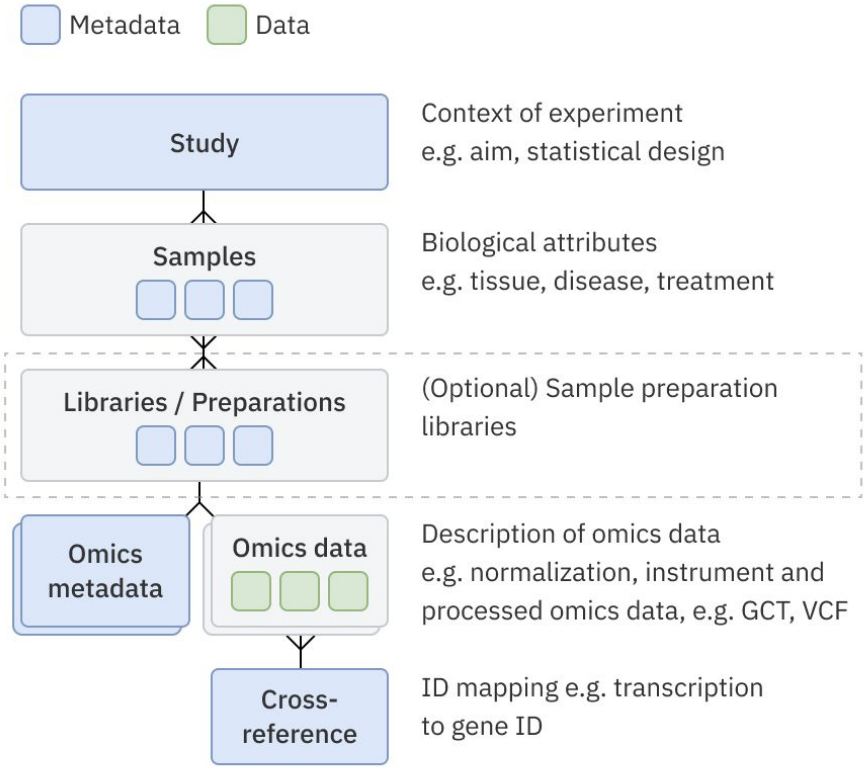
Requirements for Data Loaded to ODM

Genestack

Data Model and Supported formats



Import	GUI	Python Script
Study	Manual	TSV
Samples	Manual or TSV import	TSV
Libraries / Preparations (optional)	✗	TSV
Omics metadata	✗	TSV
Omics data	✗	GCT, VCF, FACS

Data Requirements

1. **Format.**
   All metadata should be in .tsv format.
   Expression data in .gct, variant data in .vcf.

2. **Mandatory attributes.**
   Study “Study Title” - used for displaying the name.
   Samples “Sample Source” with values and “Sample Source ID” with unique values
   Libraries “Library ID” with unique values
   Preparations “Preparation ID” with unique values

3. **No duplicated attributes**

Study Title	…
The name to be displayed	…

Sample Source	Sample Source ID	…
Source_1	Sample_1_ID	…
Source_1	Sample_2_ID	…

Library ID	Sample Source ID	…
Library_1_ID	Sample_1_ID	…
Library_2_ID	Sample_2_ID	…

Linking Attributes

Headers in the expression file should correspond to sample metadata so the system recognise what column relates to what sample.

During import user can specify which attribute from the template should be used for linking during import.

If nothing is specified the “Sample Source ID” attribute is used.

Can be any template attribute

Sample file

Name	Sample Source	Sample Source ID
Sampe_1	Lab_1	Sample_1_ID
Sample_2	Lab_1	Sample_2_ID
Sample_3	Lab_1	Sample_3_ID

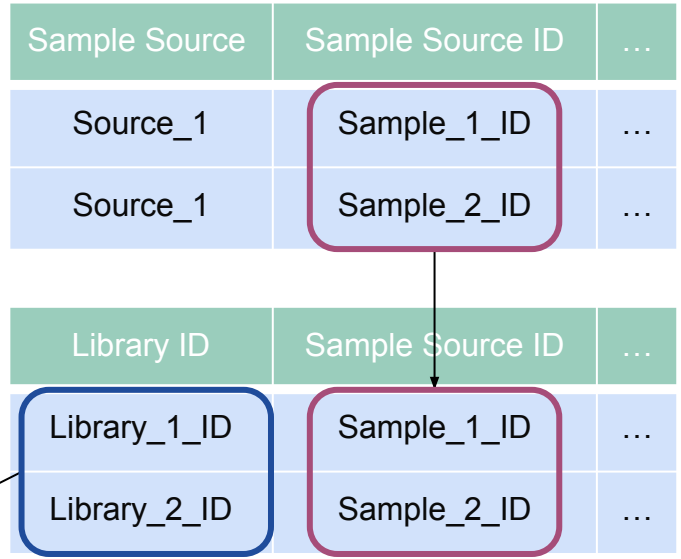
Expression file

#1.2			
24571	60		
Name	Description	Sample_1_ID	Sample_2_ID
Gene_1		2.23451	0

Linkage for Model with Libraries/Preps

For automated linking Libraries/Preparations to should have the “Sample Source ID” attribute with corresponding samples’ IDs.

Headers in the expression file should correspond to values in the “Library ID”.

Custom linking attributes are not supported.

#1.2			
24571	60		
Name	Description	Library_1_ID	Library_2_ID
Gene_1		2.23451	0