

# **GHGA Metadata: Quick Submission Guide**

version: v.0.9.0

## DATA USE CONDITIONS

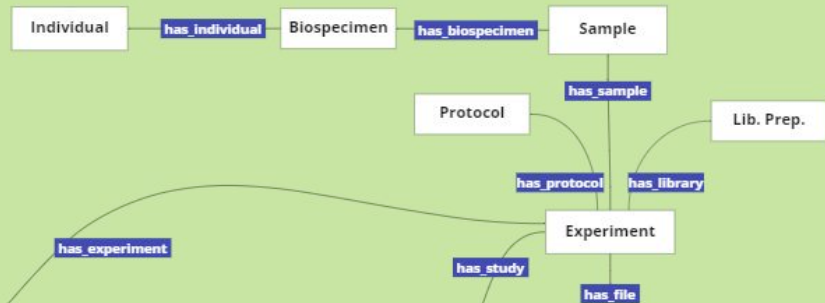


## DATASET

has\_data\_access\_policy

Dataset

## EXPERIMENT



has\_experiment

has\_project

has\_publication

has\_study

has\_file

## ANALYSIS

has\_analysis

has\_study

has\_input

has\_output

Analysis

File

# Notes: Spreadsheet, JSON, Ontology Mapping

- spreadsheet:
  - [https://github.com/ghga-de/ghga-metadata-schema/blob/main/submission-artifacts/metadata\\_submission.xlsx](https://github.com/ghga-de/ghga-metadata-schema/blob/main/submission-artifacts/metadata_submission.xlsx)
- JSON:
  - <https://github.com/ghga-de/ghga-metadata-schema/blob/main/artifacts/jsonschema/ghga.schema.json>
  - JSON and Spreadsheet are not 100% identical because the Transpiler Service fetches fields from the different spreadsheets (e.g. Sample, Core, Experiment)
- Ontology Mapping:
  - <https://docs.google.com/spreadsheets/d/1QySwEPyKIu-um8-v70yIhZQoNpdF1lurEJK3eQjweIc/edit#gid=0> (v.0.9.0 document)

# Core Submission

- Study **(mandatory)**
  - fill once for the study
  - only one study per submission supported
- Project **(optional)**
  - one row for each project
- Publication **(optional)**
  - one row for each publication

Study row does not need to be repeated if multiple Projects or Publications are submitted for the study

# File Submission

- File **(mandatory)**
  - one row for each file
  - please include ALL files you want to submit, including input and output files for analyses, phenopackets, ped-files, or other files necessary for the submission

# Experimental Submission

- Experiment **(mandatory)**
  - one experiment captures ONE library preparation and ONE sequencing protocol
  - enter a new experiment if you submit more than one library preparation and/or more than one sequencing protocol
- Library Preparation Protocol **(mandatory)**
  - one row for each protocol
- Sequencing Protocol **(mandatory)**
  - one row for each protocol

# Sample Submission

- Sample **(mandatory)**
  - one row for each sample
- Biospecimen **(optional)**
  - one row for each biospecimen
  - if you submit multiple samples, repeat the biospecimen row for each sample that belongs to that biospecimen
- Individual **(mandatory)**
  - one row for each individual
  - if you submit multiple samples and/or biospecimen, repeat the Individual row for each sample/biospecimen that belongs to that Individual

# Analysis Submission (optional)

- Analysis (optional)
  - one row per output file



# Dataset Submission

- Dataset **(mandatory)**
  - one row for each dataset
  - if you submit more than one experiment for a dataset, list experiments separated by semicolon
  - if experiment does not clearly link which file was used in which dataset, provide “submitter file alias” as semicolon-delimited list
  - if experiment does not clearly link which file was used in which dataset, provide “submitter sample alias” as semicolon-delimited list
- Data Access Committee **(mandatory)**
  - one row for each dataset. One dataset can only have one DAC
- Data Access Policy **(mandatory)**
  - one row for each dataset. One dataset can only have one DAP