

Harmonization rules and mapping to ontologies

Harmonize this values so they can be reached through queries and or filters

– BIOSAMPLE

☐ biosmple_type:

biosample_type tags > tissue, env_medium, isolation_source

biosample_type values > sample_attributes_values\$tissue, sample_attributes_values\$env_medium, sample_attributes_values\$isolation_source[-c(5,6,10)] > Map to UBERON ontology

blaf <- c("Bronchoalveolar lavage fluid", "bronchoalveolar lavage fluid(BALF)", "Human BALF sample") → "Bronchoalveolar lavage fluid" (UBERON:)

cell_culture <- c("passage") → "cell culture isolate" (UBERON:)

swab <- c("swab", "Oro-pharyngeal swab", "oropharyngeal swab") → "Oro-pharyngeal swab" (UBERON:)

unknown <- c("missing", "unknown")

☐ culture_cell:

culture_cell tags> Laboratory Host, passage_history

culture_cell values > sample_attributes_values\$Laboratory Host, sample_attributes_values\$passage_history > Map to CL ontology

vero <- "A total of 4 passages; Vero (3) at CDC followed by Vero E6 (1) at BEI Resources" → "Vero" (CL:)

vero_e6 <- c("Vero E6 cells (CRL-1586)", "A total of 4 passages; Vero (3) at CDC followed by Vero E6 (1) at BEI Resources") → "Vero E6" (CL:)

☐ collection_date: different formats e.g "02-Jan-2020", "2020-02-14" , "2020", "2020-03" (homogenize to ISO 8601 international standard)

collection_date tags> attributes_values\$collection_date

colection_date_values in formats "%d-%B-%Y", "%Y-%m-%d", "%Y-%m", "%Y"

–HOST/INDIVIDUAL

☐ host_taxon_id tags> sample_attributes_values\$host, sample_attributes_values

\$env_broad_scale, sample_attributes_values\$host_description

host_species_name values>

human <- c("Homo sapiens", "Human", "Isolated from clinical sample from the first case of nCov in US from Washington State") → "9606"

☐ geo_origin:

geo_origin tags > geo_loc_name, country, at_lon

geo_origin_values > sample_attributes_values\$geo_loc_name, sample_attributes_values

\$country, sample_attributes_values\$at_lon

(different formats) e.g "USA:WI:Madison"/ "USA: CA, San Diego County"/ "30.52 N 114.31

E" > harmonise, map to GAZ ontology (geographic origin in default schema)