

Metadata fields from 50 Illumina files to feed virus beacon schema v1 table

*NOTE: some values need to be harmonized among datasets (see Harmonization rules and mapping ontologies doc)

–VARIANT IN SAMPLE

☐ biosample_id: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$SAMPLE\$IDENTIFIERS\$PRIMARY_ID¹

☐ variant_file_id: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$RUN_SET\$RUN\$IDENTIFIERS\$PRIMARY_ID¹

• INFO

☐ study_info:

☐ study_id: (study accession): xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$STUDY\$IDENTIFIERS\$PRIMARY_ID¹

☐ study_title: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$STUDY\$DESCRIPTOR\$STUDY_TITLE¹

☐ study_ref: (article PUMED ID)

☐ experiment_info

☐ exp_id: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$EXPERIMENT\$IDENTIFIERS\$PRIMARY_ID

☐ exp_lib_strategy: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$EXPERIMENT\$DESIGN\$LIBRARY_DESCRIPTOR\$LIBRARY_STRATEGY

☐ exp_lib_source: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$EXPERIMENT\$DESIGN\$LIBRARY_DESCRIPTOR\$LIBRARY_SOURCE¹

☐ exp_lib_selection: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$EXPERIMENT\$DESIGN\$LIBRARY_DESCRIPTOR\$LIBRARY_SELECTION¹

☐ exp_lib_layout: names(xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$EXPERIMENT\$DESIGN\$LIBRARY_DESCRIPTOR\$LIBRARY_LAYOUT)

☐ exp_platform: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$EXPERIMENT\$PLATFORM\$ILLUMINA\$INSTRUMENT_MODEL¹

–BIOSAMPLE

☐ biosample_id: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$SAMPLE\$IDENTIFIERS\$PRIMARY_ID¹

☐ biosample_alt_id: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$SAMPLE

\$IDENTIFIERS\$EXTERNAL_ID¹

- ☐ biosample_type: sample_attributes_values\$tissue, attributes_values\$isolation_source
- ☐ culture_cell: sample_attributes_values\$Laboratory Host, attributes_values
\$passage_history > Map to CL ontology (NULL or none if not culture)
- ☐ culture_passage_history: attributes_values\$passage_history e.g "Original (not
passaged)" (NULL or none if not culture)
- ☐ collection_date: attributes_values\$collection_date
- ☐ study_ref (article PUMED ID): sample_attributes_values\$link_addit_analys

–HOST/INDIVIDUAL

- ☐ host_taxon_id: sample_attributes_values\$host, sample_attributes_values
\$env_broad_scale, sample_attributes_values\$host_description, sample_attributes_values
\$country
- ☐ host_age: sample_attributes_values\$host_age, sample_attributes_values\$age
- ☐ host_sex: sample_attributes_values\$host_sex, sample_attributes_values\$sex
- ☐ geo_origin: sample_attributes_values\$geo_loc_name, sample_attributes_values\$lat_lon
- ☐ host_disease: sample_attributes_values\$host_disease
- ☐ host_comorbidities (diseases in default schema): not present in initial 50 files
- ☐ host_disease_outcome: sample_attributes_values\$host_disease_outcome
- ☐ info
 - ☐ study_ref (article PUMED ID): sample_attributes_values\$link_addit_analys

– VIRUS

- ☐ taxon_id: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$SAMPLE
\$SAMPLE_NAME\$TAXON_ID¹
- ☐ taxon_name: xml\$EXPERIMENT_PACKAGE_SET\$EXPERIMENT_PACKAGE\$SAMPLE
\$SAMPLE_NAME\$SCIENTIFIC_NAME¹
- ☐ strain_id:
- ☐ strain_name: sample_attributes_values\$strain