

virus_beacon_schema_v1_to_generic

–VARIANT BASIC (basic beacon variant schema: from VCF)

- ☐ taxon_id (taxon of sequenced species)
- ☐ ref_assembly_id_version:
- ☐ start_nucleotide
- ☐ end_nucleotide
- ☐ ref_sequence
- ☐ alt_sequence
- ☐ variantType: SNV, indel, CNV, structural variant

–VARIANT ANNOTATION

- ☐ variant_id: (external reference if it exists)
- ☒ ~~variant_type: e.g "del" (belongs in beacon variant basic)~~
- ☐ variant effect: e.g "missense variant" → molecularConsequence
- ☐ genomic_region: categorical, from virus genomic annotation in VIRUS: annotation (SARS-CoV2: 5UTR, ORF1ab, S, ORF3a, Intergenic, E, M, ORF6, ORF7a, ORF8, N, ORF10, 3UTR) → genomicRegionClass and featureId
- ☒ ~~functional_region: categorical, from functional annotation file VIRUS: annotation e.g "HVR", "RBD", "RNA modification site" or manual entering by user.~~

–VARIANT IN SAMPLE

- ☐ biosample_id: (external ref) e.g "SRS6007144" → Variant in Sample biosampleId
- ☐ host_id: (external ref if it exists) → Variant in Sample individualId
- ☐ variant_frequency_dataset: from vcf → Variant in Sample variantFrequency
- ☐ info
 - ☒ study_info:
 - ☒ study_id: (study accession): e.g "SRP242226"
 - ☒ study_ref: (article PUMED ID or URL)
 - ☐ experiment_info
 - ☒ variant_file_id: (accession if external ref or internal if we run pipeline) → (to info?)
 - ☒ sequence_file_id: (run accession) e.g "SRR10903401"
 - ☒ exp_id (experiment accession): e.g "SRX7571571"
 - ☒ exp_title: e.g "Total RNA sequencing of BALF (human reads removed)"

- ☐ exp_lib_strategy: ("RNA-Seq", "WGS", "AMPLICON", "Targeted-Capture") → Variant in Sample libraryStrategy
- ☒ exp_lib_source: (~~"METATRANSCRIPTOMIC", "METAGENOMIC", "GENOMIC", "VIRAL-RNA"~~)
- ☒ exp_lib_selection: (~~"RANDOM", "RT-PCR", "RANDOM-PCR", "unspecified", "PCR", "cDNA"~~)
- ☒ exp_lib_layout: (~~"PAIRED" "SINGLE"~~)
- ☐ exp_platform: ("Illumina", "Nanopore") → Variant in Sample seqPlatform
- ☒ exp_platform_model: (~~"Illumina-MiSeq", "Illumina-MiniSeq", "Illumina-HiSeq-2500", "NextSeq-500", "NextSeq-550", "Illumina-iSeq-100", "GridION"~~)
- ☐ variant_caller: (from VCF or Galaxy or Master of Pores pipelines) → Variant in Sample variantCaller

–BIOSAMPLE

- ☐ biosample_id: (external ref) e.g "SRS6007144"
- ☐ biosample_alt_id: (external ref) e.g "SAMN13872787"
- ☐ individual_id
- ☐ collection_date: e.g "2020-02-14"
- ☐ biosample_type: (sample type/source) e.g "Bronchoalveolar lavage fluid" or "Cellular passage" → sampleOrigin
- ☐ procedure: → obtentionProcedure
 - ☒ culture_cell: e.g: "Vero-E6 cells (CRL 1586)" (NULL or none if not culture)
 - ☒ culture_passage_history e.g "Original (not passaged)" (NULL or none if not culture)
- ☐ info
 - out
 - ☐ biosample_ref_material: e.g "BEI Resources catalog NR-52281 (lot 70033135)"

– ~~HOST~~ INDIVIDUAL (ON THIS 1 ENCOUNTER)

- ☐ individual_id: (external ref)
- ☐ host_taxon_id: e.g "9606" ("Homo sapiens") → where?
- ☐ host_age: e.g "21" → Individual age
- ☐ host_sex: "female", "male" (sex in default schema) → Individual sex
- ☐ geo_origin: e.g "USA:WI:Madison" → Individual geographicOrigin
- ☐ disease: (relevant virus-related diseases) e.g "COVID 19 pneumonia" → Individual diseasesDisease

- ☐ disease_stage: e.g "acute" → Individual diseasesStage
- ☐ comorbidities: (underlying chronic diseases, format as individualDiseases from default schema): e.g ICD10 for "diabetes mellitus type II" → Individual diseasesDisease and diseasesStage where they are set stage: chronic
- ☒ disease_course: categorical "~~asymptomatic~~", "~~mild~~", "~~severe~~"
- ☒ disease_outcome: e.g "~~resolution/discharge~~", "~~fatal~~"
- ☐ info
 - ☐ → out

~~VIRUS~~ (Should there be one generic for organism/entity which sequence data belongs to to include their relevant phenotypic features - in the form of whatever phenotypic feature ontology for any species, or one specific to Genus , or being removed altogether?)

- ☐ taxon_id: e.g "433733" → Variant Basic taxonId
- ☒ taxon_name: e.g "~~Severe acute respiratory syndrome coronavirus 2~~"
- ☒ strain_id:
- ☒ strain_name: e.g "~~2019 nCoV/USA WI1/2020~~" → info?
- ☒ annotation
 - ☒ genomic annotation: file url
 - ☒ functional annotation: file url