Beacon v.2 schema

Dataset

datasetId CATEGORICAL VALUE Dataset reference ID

datasetName CATEGORICAL VALUE Dataset name. Reference to source. e.g DECIPHER, DisGenNET

datasetype CATEGORICAL VALUE Dataset type: variant-level (aggregated) or case-level

Variant Identification

variantId ALPHANUMERIC VALUE ID referencing the variant in beacon (internal ID)

assemblyId CATEGORICAL VALUE Genomic assembly accession and version as RefSqeq assembly accessions (e.g "GCG_000001405.39"). Alternatively, an assembly name or synonym such as UCSC Genome Browser assembly (e.g "hg38") or Genome Reference consortium Human (e.g GRCh38.p13") names can be given as long as they are accompanied with their versions.

refseqId CATEGORICAL VALUE Reference sequence Refseq ID and version for genomic contiguous in which variant query coordinates are given, e.g "NC_000009" for human chromosome 9. Alternatively, names, synonymous or aliases are accepted eg. "Chr9" when **assemblyId** is given. For organism with single scaffold the full length reference sequence Refseq IDs can be given here as an alternative to the assembly Id and version in **assemblyId**, e.g "NC_045512.2" for SARS-CoV2 full-length genome reference sequence.

start NUMERIC VALUE Start position of variant

end NUMERIC VALUE End position of variant

ref ALPHANUMERIC VALUE Reference sequence in start-end coordinates

alt ALPHANUMERIC VALUE Alternative sequence in start-end coordinates

variantType CATEGORICAL VALUE (ONTOLOGY LABEL)

Variant Annotation

variantId ALPHANUMERIC VALUE ID referencing the variant in beacon (internal ID)

variantAlternativeId (LIST OF) ALPHANUMERIC VALUE(S) Cross-referencing ID(s) (CURIE(s)) for the variant in the original databases or variant-level (aggregated) databases for previously described variants (e.g. clinVarId, ClinGen, COSMIC), e.g : "VCV000055583.1", "rs80356868", "CA003602"

genomicHGVSId ALPHANUMERIC VALUE HGVSId descriptor at genomic level (recommended, referred to genome assembly defined in Variant Basic), e.g "NC_000017.10:g.41199678C>A"

transcriptHGVSId LIST OF ALPHANUMERIC VALUE(S) HGVSId descriptor at transcript level : "NC_000023.10(NM_004006.2):c.357+1G

proteinHGVSId LIST OF ALPHANUMERIC VALUE(S) (List of) HGVSId descriptor(s) at protein level (for protein-altering variants), e.g "NP_009225.1:p.Glu1817Ter" or LRG_199p1:p.Val25Gly (preferred)

genomicRegion CATEGORICAL VALUE (ONTOLOGY LABEL) (List of) Classification(s) of the variant according to the genomic region affected (all that apply). Value from Sequence Ontology (SO) (SO:TBD), e.g "intergenic", "5UTR", "3UTR", "coding"

genomicFeatures Genomic feature(s) affected by the variant. (List of:)

class CATEGORICAL VALUE (ONTOLOGY LABEL) Class of genomic region affected by the variant eg "gene" "protein coding transcript", "untranslated region", "non-coding transcript"

featureID (ALPHANUMERIC VALUE) ID /accession/name of genomic region affected by the variant, matching class in **class**, e.g "TP53", "GeneID:43740578"

annotationToolVersion ALPHANUMERIC VALUE Tool used for annotation and prediction of variant effects e.g "SnpEffVersion=4.3t (build 2017-11-24 1018)"

molecularEffect ALPHANUMERIC VALUE (List of) Predicted effect at nucleotide level eg. "STOP_GAINED"

molecularConsequence CATEGORICAL VALUE (ONTOLOGY LABEL) (List of) Predicted effect at protein level for protein affecting variants eg. "nonsense" , "missense"

aminoacidChange CATEGORICAL VALUE (ONTOLOGY LABEL) (List of) Change(s) at aminoacid level for protein affecting variants eg. "V304*"

Subject

This object contains info related to the subject from where the variants are found in a study. It includes taxon id and any other relevant information about it, including maybe links or id of genome assemblies and ref seqs associated to this species that are available in the Beacon and that are used for variant identification (location)

subjectId ALPHANUMERIC VALUE Reference ID of subject (external accession or internal ID)

taxonId CATEGORICAL VALUE (ONTOLOGY LABEL) Reference taxon ID for subject organism i.e human, animal or plant, etc.

sex CATEGORICAL VALUE (ONTOLOGY LABEL) Sex of subject. Value from NCIT General Qualifier ontology (NCIT:C27993): "UNKNOWN" (not assessed or not available) (NCIT:C17998), "FEMALE" (NCIT:C46113), "MALE", (NCIT:C46112) or "OTHER SEX" (NCIT:C45908)

ethnicity CATEGORICAL VALUE (ONTOLOGY LABEL) Ethnic background of subject. Value from NCIT Race ontology (NCIT:C17049). e.g "Latin American" (NCIT:C126531)

geographicOrigin CATEGORICAL VALUE (ONTOLOGY LABEL) Subject's country or region of origin (birthplace or residence place regardless of ethnic origin). Value from GAZ Geographic Location ontology (GAZ:00000448), e.g. "United States of America" (GAZ:00002459)

phenotypicFeatures Phenotypic feature(s) observed in the subject, defined by phenotype, age of onset and level/ severity. (List of:)

phenotypeId CATEGORICAL VALUE (ONTOLOGY LABEL) Phenotypic feature observed. Value from Human Phenotype Ontology (HPO) or other phenotype ontology

dateOfOnset ALPHANUMERIC VALUE Date of onset/observation of phenotype, in (ISO8601 duration format)

ageOfOnset Subject's age at onset/observation of phenotype

age ALPHANUMERIC Age, in (ISO8601 duration format)

ageGroup CATEGORICAL VALUE (ONTOLOGY LABEL) Age group value, from NCIT Age Group ontology, e.g. "NCIT:C27954" (Adolescent)

level/severity CATEGORICAL VALUE (ONTOLOGY LABEL) Level/severity when and as applicable to phenotype observed. Value from TBD, e.g "mild"

diseases Disease(s) been diagnosed to the subject, defined by disease ID, age of onset, stage, level/severity, outcome and the presence of family history. (List of:)

diseaseId CATEGORICAL VALUE (DISEASE CODE/ONTOLOGY LABEL Disease ID. Value from ICD10 disease codes or ontology terms from disease ontologies such as HPO, OMIM, Orphanet, MONDO. e.g. "lactose intolerance" (HP:0004789, ICD10CM:E73)

dateOfOnset ALPHANUMERIC VALUE (ISO8601 DURATION FORMAT) Date of onset/diagnosis of disease

ageOfOnset Subject's age at onset/ diagnosis of disease

age ALPHANUMERIC Age, in (ISO8601 duration format)

ageGroup CATEGORICAL VALUE (ONTOLOGY LABEL) Age group value, from NCIT Age Group ontology, e.g. "NCIT:C27954" (Adolescent)

stage CATEGORICAL VALUE (ONTOLOGY LABEL) from Ontology for General Medical Science or Disease Stage Qualifier ontology (NCIT:C28108) . e.g. "acute onset" (OGMS:0000119)

level/severity CATEGORICAL VALUE (ONTOLOGY LABEL) Level/severity when and as applicable to disease course.Value from TBD, e.g "severe"

outcome CATEGORICAL VALUE (ONTOLOGY LABEL) Outcome of passed acute diseases. Value from TBD, eg. "fatal"

familyHistory BOOLEAN indicating determined or self-reported presence of family history of the disease

treatments Treatment(s) been prescribed/administered to subject, defined by treatment ID), date and age of onset, dose, schedule and duration. (List of:)

treatmentId CATEGORICAL VALUE (ONTOLOGY LABEL) Treatment ID. Value from TBD

dateAtOnset ALPHANUMERIC VALUE Date of the beginning of treatment, in (ISO8601 duration format)

ageAtOnset Subject's age at the beginning of treatment

age ALPHANUMERIC Age, in (ISO8601 duration format)

ageGroup CATEGORICAL VALUE (ONTOLOGY LABEL) Age group value, from NCIT Age Group ontology, e.g. "NCIT:C27954" (Adolescent)

dose NUMERIC Treatment dose

units ALPHANUMERIC Treatment dose units

schedule CATEGORICAL VALUE (ONTOLOGY LABEL) Treatment schedule. Value from TBD, e.g "weekly"

duration ALPHANUMERIC VALUE Treatment duration, in (ISO8601 duration format)

interventions Intervention(s) been practiced on subject, defined by treatment ID), date and age of onset, dose, schedule and duration. (List of:)

interventionId CATEGORICAL VALUE (ONTOLOGY LABEL) Intervention ID. Value from TBD

date ALPHANUMERIC VALUE Date of intervention, in (ISO8601 duration format)

ageAtIntervention Subject's age at the date of intervention in age or age range

age ALPHANUMERIC Age, in (ISO8601 duration format)

ageGroup CATEGORICAL VALUE (ONTOLOGY LABEL) Age group value, from NCIT Age Group ontology, e.g. "NCIT:C27954" (Adolescent)

pedigrees list of:

pedigreeID ALPHANUMERIC VALUE ID referencing pedigree

pedigreeRole CATEGORICAL VALUE (ONTOLOGY LABEL) Pedigree role, defined as relationship to proband. Value from HL7 code for family relationship or Relationship to Proband ontology (ERO:0002112) . e.g "self" (ERO:002036), "identical twin relationship" (ERO:0002041)

numIndTested NUMERIC VALUE

Biosample

biosampleId ALPHANUMERIC VALUE ID referencing the biosample (external accession)

subjectId ref to Subject's subjectId

description FREE TEXT Any relevant info about the biosample that does not fit in any field in the schema

biosampleStatus CATEGORICAL VALUE (ONTOLOGY LABEL) from Experimental Factor Ontology (EFO) Material Sample ontology (OBI:0000747) Classification of the sample in "abnormal sample "(EFO:0009655) or "reference sample" (EFO:0009654)

collectionDate ALPHANUMERIC VALUE(ISO8601 DURATION FORMAT) Date of biosample collection

subjectAgeAtCollection ALPHANUMERIC VALUE (ISO8601 DURATION FORMAT) Subject's age at the time of sample collection

sampleOriginType CATEGORICAL VALUE (ONTOLOGY LABEL) Category of sample origin e.g "organism primary tissue", "organism xenograft", "organism-derived fluid", "cell culture", "environmental sample"

sampleOriginDetail CATEGORICAL VALUE (ONTOLOGY LABEL) from Uber-anatomy ontology (UBERON) or BRENDA tissue / enzyme source (BTO) Specific instance of sample origin matching the category set in sampleOriginType e.g "HEK293T", "nasopharynx"

obtentionProcedure CATEGORICAL VALUE (ONTOLOGY LABEL) Ontology ID from Intervention or Procedure NCIT ontology. e.g. "biopsy" (NCIT:C15189)

cancerFeatures Values specifying cancer-specific features, including progression and tumor grade

> **tumorProgression** CATEGORICAL VALUE (ONTOLOGY LABEL) from Neoplasm by Special Category ontology (NCIT:C7062). Tumor progression category indicating primary, metastatic or recurrent progression e.g "Primary Malignant Neoplasm" (NCIT:C84509)
>
> **tumorGrade** CATEGORICAL VALUE (ONTOLOGY ID) from Tumor Grading Characteristic ontology (Mondo Dis- ease Ontology MONDO:0024488) General tumor grading

Run

runId ALPHANUMERIC VALUE Internal or external accession e.g "SRR10903401"

biosampleId ALPHANUMERIC VALUE Reference to sample

variantId TIMESTAMP Date at which run was performed, in (ISO8601 duration format)

librarySource CATEGORICAL VALUE (ONTOLOGY LABEL) Sequencing library source e.g "Metagenomic", "Viral RNA"

libraryStrategy CATEGORICAL VALUE (ONTOLOGY LABEL) Sequencing library strategy e.g "WGS"

librarySelection CATEGORICAL VALUE (ONTOLOGY LABEL) Selection method for sequencing library preparation e.g "RANDOM", "RT-PCR"

libraryLayout CATEGORICAL VALUE (ONTOLOGY LABEL) Sequencing library layout e.g "PAIRED", "SINGLE"

platform CATEGORICAL VALUE (ONTOLOGY LABEL) Sequencing platform group e.g "Illumina", "Nanopore"

platformModel CATEGORICAL VALUE (ONTOLOGY LABEL) Sequencing platform model e.g "Illumina MiSeq", "GridION"

Analysis

runId ALPHANUMERIC VALUE Internal or external accession e.g "SRR10903401"

variantId TIMESTAMP Date at which analysis was performed, in (ISO8601 duration format)

pipelineName CATEGORICAL VALUE Analysis pipeline and version if a standardized pipeline was used

pipelineRef Link to Analysis pipeline resource

aligner CATEGORICAL VALUE (ONTOLOGY LABEL) Mapping/Alignment software e.g bwa

variantCaller CATEGORICAL VALUE (ONTOLOGY LABEL) Variant calling software/ pipeline e.g "GATK vxxx"

Variant in Sample

variantId ALPHANUMERIC VALUE

analysisId ref Run runId

subjectId ref Subject's subjectId

variantFrequency NUMERIC VALUE Variant frequency in sample, as in AF field in VCF for case-level datasets. Frequency in dataset for aggregated variant-level datasets.

zigosity CATEGORICAL VALUE (ONTOLOGY LABEL) Zigosity in which variant is present in the sample from the Zigosity Ontology (GENO:0000133) , e.g "heterozygous" (GENO:0000135)

alleleOrigin CATEGORICAL VALUE (ONTOLOGY LABEL) Allele origin of variant in sample from the Variant Origin (SO:0001762). Categories are "somatic variant", "germline variant", "maternal variant", "paternal variant", "de novo variant", "pedigree specific variant", "population specific variant". Corresponds to Variant Inheritance in FHIR.

phenotypicEffect CATEGORICAL VALUE (ONTOLOGY LABEL) Annotated effect on disease. list of:

- **phenotypeId** CATEGORICAL VALUE (ONTOLOGY LABEL) Descriptor of phenotype found associated in this study
- **phenotypeEffect** CATEGORICAL VALUE (ONTOLOGY LABEL) Phenotypic effect classification determined in this study
- **evidenceType** CATEGORICAL VALUE (ONTOLOGY LABEL) Type of evidence supporting variant-phenotype association from the Evidence & Conclusion Ontology (ECO) e.g "experimental evidence"

clinicalRelevance CATEGORICAL VALUE (ONTOLOGY LABEL) Annotated effect on disease. list of:

- **diseaseId** CATEGORICAL VALUE (ONTOLOGY LABEL) Descriptor of disease associated
- **clinicalEffect** CATEGORICAL VALUE (ONTOLOGY LABEL) Clinical effect classification
- **evidenceType** CATEGORICAL VALUE (ONTOLOGY LABEL) Type of evidence supporting variant-disease association from the Evidence & Conclusion Ontology (ECO)

Variant Interpretation

variantId ALPHANUMERIC VALUE ID referencing the variant in beacon (internal ID)

datasetId ALPHANUMERIC VALUE ID referencing the dataset from variant interpretation

phenotypicEffect (List of) Annotated effects on any phenotypic feature other than a disease. (List of:)

- **phenotypeId** CATEGORICAL VALUE (ONTOLOGY LABEL) Descriptor of phenotype associated
- **phenotypeEffect** CATEGORICAL VALUE (ONTOLOGY LABEL) Phenotypic effect classification

alleleOrigin CATEGORICAL VALUE(S) (ONTOLOGY LABEL) (List of) Annotation(s) on allele origins in which the variant has been found in association to condition. Categories are "somatic variant", "germline variant", "maternal variant", "paternal variant", "de novo variant", "pedigree specific variant", "population specific variant". Corresponds to Variant Inheritance in FHIR.

references (List of) PMID(s)

clinicalRelevance Annotated effect on disease. (List of:)

diseaseId CATEGORICAL VALUE (ONTOLOGY LABEL) Descriptor of disease associated

clinicalEffect CATEGORICAL VALUE (ONTOLOGY LABEL) Clinical effect classification

alleleOrigin CATEGORICAL VALUE(S) (ONTOLOGY LABEL) (List of) Annotation(s) on allele origins in which the variant has been in association to condition. Categories are "somatic variant", "germline variant", "maternal variant", "paternal variant", "de novo variant", "pedigree specific variant", "population specific variant". Corresponds to Variant Inheritance in FHIR.

references (List of) PMID(s)

Interactor

This is an organism/agent whose metadata/ phenotypic data is collected in association with the Subject, but which is not sequenced itself. It accounts for 'extended phenotype' of variants in other organisms/agents than the one harboring them.

relationType CATEGORICAL VALUE (ONTOLOGY LABEL) Type of relation with Subject e.g "host", "pathogen", "commensal", etc

[...] All the rest of objects from Subject