Spec for Individual, Biosample, Variant Annotation and Variant in Individual schemas Beacon 2.0

taxonID

Specifications and examples for the format and content of fields in Biosample, Individual, Variant Annotation and Variant in Individual schemas for Beacon 2.0.

Individual

The Individual schema is part of a Data Beacon 2.0 schema, i.e, it applies to variants whose source is an Individual, as opposed to those variants in Evidence Beacon 2.0 schema where the variant source is a Knowledge Base (DisGeNET like schema).

datasetID ID referencing the dataset of provenance in the database

individualId ID referencing the individual in the database

~~**age** Definition of age of individual at baseline as age and/or age group~~

~~age Value indicating the age of the individual at the time of collection in the ISO8601 duration format P[n]Y[n]M[n]DT[n]H[n]M[n]S. e.g "P32Y6M1D"~~

~~ageGroup Categorical value from NCIT Age Group ontology classifying the individuals in age groups. e.g. "NCIT:C27954" (Adolescent)~~

sex Categorical 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 from NCIT Race ontology (NCIT:C17049). e.g Latin American (NCIT:C126531)

geographicOrigin Categorical value from GAZ Geographic Location ontology (GAZ:00000448) describing the individual´s country or region of origin (birthplace or residence place regardless of ethnic origin). e.g. "GAZ:00002459" (United States of America)

diseases (List of) disease(s) been diagnosed to the individual, defined by disease ICD10 code, optionally other disease ontology ID(s), age of onset, stage and the presence of family history

disease Disease ICD10 code or ontology ID from a disease ontology such as HPO, OMIM, Orphanet, MONDO. e.g. "ICD10CM:E73" or "HP:0004789"

ageOfOnset age format from Individual schema.

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

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

pedigrees (List of) pedigree studie(s) the individual has taken part in

pedigreeId ID referencing the pedigree in the database

disease Ontology ID indicating the disease being the focus of the pedigree

pedigreeRole Value from HL7 code for family relationship or Relationship to Proband ontology (ERO:0002112). e.g "ERO:0002041" (identical twin relationship)

numberOfIndividualsTested Numeric indicating number of family members, including the proband, studied in the pedigree

info

Biosample

The Biosample schema is part of a Data Beacon 2.0 schema, i.e, it applies to variants whose source is an Individual, as opposed to those variants in Evidence Beacon 2.0 schema where the variant source is a Knowledge Base (DisGeNET like schema).

individualId ID referencing the individual of provenance in the database

bioSampleId ID referencing the biosample in the database

description Free text describing any relevant info about the biosample that does not fit in any field in the schema

biosampleStatus Categorical value from Experimental Factor Ontology (EFO) Material Sample ontology (OBI:0000747) classifying the sample in abnormal sample (EFO:0009655) or reference sample (EFO:0009654). e.g. "EFO:0009655" (abnormal sample)

individualAgeAtCollection age format from Individual schema

sampleOrigin Values of ontology IDs specifying the origin of the biosample in organ, tissue and cell type/cell line. For samples with origin in an established cell line, the cell line identifier should be specified in the cell type field

biosampleType
sampleOriginType

~~organ~~ ~~Ontology ID from Uber anatomy ontology (UBERON) or BRENDA tissue / enzyme source (BTO) ontologies identifying the source organ of the biosample. e.g. "UBERON:0002107" (liver)~~

~~tissue~~ ~~Ontology ID from Uber anatomy ontology (UBERON) or BRENDA tissue / enzyme source (BTO) ontologies identifying the source tissue of the biosample. e.g. "UBERON:0001281" (hepatic sinusoid)~~

~~cellType~~ ~~Ontology ID from BRENDA tissue / enzyme source (BTO) or Cell Ontology (CL) ontologies identifying the source cell type or cell line origin of the biosample. e.g. "CL:0000091" (Kupffer cell)~~

obtentionProcedure Ontology ID from Intervention or Procedure NCIT ontology. e.g. "NCIT:C15189" (Biopsy)

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

~~tumorProgression Value from tumor progression indicating primary, metastatic or recurrent progression categories Neoplasm by Special Category ontology (NCIT:C7062). e.g "NCIT:C84509" (Primary Malignant Neoplasm)~~

~~tumorGrade~~ ~~Value from general tumor grading Tumor Grading Characteristic ontology (Mondo Disease Ontology MONDO:0024488). A modifier that can be applied to a tumor class describing abnormal tumor histology or morphology. It is an indicator of how quickly a tumor is likely to grow and spread. Grading systems differ depending on the type of cancer. e.g. "MONDO:0024492" (tumor grade 2, general grading system)~~

info

Variant annotation

The Variant Annotation schema is common to both an Evidence and a Data Beacon.

variantId ID referencing the variant in database

alternativeIds (List of) Cross-referencing ID(s) (CURIE) for previously described variants (e.g. clinVarId, ClinGen, COSMIC), e.g "alternativeIds": ["VCV000055583.1", "rs80356868", "CA003602"]

genomicHGVSId HGVSId descriptor at genomic level (recommended, referred to genome assembly defined in Variant Representation schema), e.g "genomicHGVSId": "NC_000017.10:g.41199678C¿A"

proteinHGVSIds (List of) HGVSId descriptor(s) at protein level (for protein-altering variants), e.g "NP_009225.1:p.Glu1817Ter"

molecularConsequence Categorical value from Sequence Variant ontology (SO:0001060) describing the molecular consequence of the variant such as missense variant and frameshift truncation variant for protein-altering variants, e.g "SO:0001583" (missense variant)

~~**variantGeneRelationship**~~ : Categorical value classifying the variant according to the broadness of the variant effect in terms of ~~genes~~: intergenic, 5UTR, 3UTR, single-gene (exonic, intronic), in overlapping genes (exonic, intronic), spanning multiple genes, multiple genes, e.g "single gene"

featureId

~~**geneId**~~ : ~~(List of) HGNC ID(s) for gene(s) affected by the variant, e.g "HGNC:8157"~~

~~**transcriptIds** (List of) ENSMEBL ID(s) for transcript(s) affected by the variant, e.g "ENST00000277010.9"~~

variantPhenotypicEffect

~~**clinicalRelevance**~~ : (List of) description elements of the ~~clinical relevance~~ described for the variant in relation to specific ~~diseases~~ including a variant classification, the disease identifier, and references

phenotype

phenotype

~~variantClassification~~ Categorical value ~~(benign, likely bening, pathogenic, likely pathogenic or unknown)~~

~~diseaseId diseaseId format from Disease object from Individual schema, e.g "MONDO:0003582" (hereditary breast ovarian cancer syndrome)~~

references (List of) PUBMED ID(s) of studie(s) describing the variant-disease association

, e.g "PMID:27153395", "PMID:27616075"

alleleOrigin (List of) Categorical value(s) from Variant Origin ontology (SO:0001762) (germline variant (SO:0001778), de novo variant (SO:0001781), somatic variant (SO:0001777)), indicating the allele origin(s) in which the variant has been found, e.g "SO:0001781" (de novo variant)

info

Variant in sample

~~Variant in Individual~~

The Variant in Individual schema is part of the Evidence Beacon schema.

variantId ID

biosampleID

individualId ID referencing individual of provenance

variantFrequency

~~**timestamp** Time of variant publication~~

zygosity Categorical value from Zigosity Ontology (GENO:0000133) indicating the zigosity in which the variant has been found in individual, e.g "GENO:0000135" (heterozygous)

alleleOrigin Categorical value from Variant Origin ontology (SO:0001762), e.g "SO:0001777" (somatic variant)

~~**clinicalRelevance** Clinical relevance as experimentally determined in current study: non pathogenic, pathogenic, likely pathogenic, unknown, evidence type instead of reference, e.g "pathogenic"~~

~~variantClassification Categorical value defining variant´s clinical relevance (non pathogenic, pathogenic, likely pathogenic or unknown) as experimentally determined in current study.~~

Text

~~disease Disease ID of associated disease (diseaseId format from Disease schema), e.g "HP:0000717"~~

info

libraryStrategy

seqPlatform

variantCaller