

BiologicalEntity specification vr. 0.0.3

Bioschemas specification describing BiologicalEntity in the life-science

Description. The BiologicalEntity aims to encompass as much as possible biological types such as “samples”, “protein”, “protein annotation”, “protein structure”, “phenotype” and so on. Most of the properties here will be optional as the specific profiles, i.e., a tailored usage of BiologicalEntity, will provide stronger requirements whenever needed.

Most of the new properties belong to BiologicalEntity but some required changes at a different level in schema.org. That is why we also have some new properties for CreativeWork and Thing.

Supporting information

- [Use cases](#)
- [Mapping/Crosswalk](#)
- [Examples](#)
- [Graph](#)

Properties

Based on schema.org [CreativeWork]

Property	Expected Type	Description	CN	MG	CV
citation	CreativeWork or URL	A citation or reference to a creative work, such as a publication, web page, scholarly article, etc.	many	R	no
dateCreated	Date or DateTime	The date on which the CreativeWork/BiologicalEntity was created or the item was added to a DataFeed/Dataset/DataRepository.	one	O	no
dateModified	Date or DateTime	The date on which the CreativeWork/BiologicalEntity was most recently modified or when the item's entry was modified within a DataFeed/Dataset/DataRepository.	one	O	no
datePublished	Date	Date of first broadcast/publication.	one	O	no
hasPart	BiologicalEntity	Indicates a CreativeWork/BiologicalEntity that is (in some sense) a part of this CreativeWork/BiologicalEntity. Inverse property: isPartOf.	many	O	no
isBasedOn	CreativeWork or URL or	A resource that was used in the creation of this resource. This term can be repeated for multiple	many	O	no

	BiologicalEntity	sources. For example, http://example.com/great-multiplication-intro.html . Supersedes isBasedOnUrl.			
isPartOf	BiologicalEntity	Indicates a CreativeWork/BiologicalEntity that this CreativeWork/BiologicalEntity is (in some sense) part of. Inverse property: hasPart.	many	0	no

New properties for [CreativeWork]

Property	Expected Type	Description	CN	MG	CV
isBasisFor	CreativeWork or URL or BiologicalEntity	A resource for which this resource has been used for the creation of the former. Inverse property: isBasedOn	many	0	no

Based on schema.org [Event]

Property	Expected Type	Description	CN	MG	CV
location	Place	Position where this entity is located or originates from (e.g. an entity from Polynesia islands or an Anatomical location. If multiple locations with multiple purposes (collection, storage) should be modelled, please use the additionalProperty from Place to specify this.	many	0	no

Based on schema.org [Dataset]

Property	Expected Type	Description	CN	MG	CV
distribution	DataDownload	A downloadable form of this dataset, at a specific location, in a specific format.	many	0	no
measurementTechnique	Text or URL	A technique or technology used in a Dataset (or DataDownload, DataCatalog), corresponding to the method used for measuring the corresponding variable(s) (described using variableMeasured). This is oriented towards scientific and scholarly dataset publication but may have broader applicability; it is not intended as a full representation of measurement, but rather as a high level summary for dataset discovery. For example, if variableMeasured is: molecule concentration, measurementTechnique could be: "mass spectrometry" or "nmr spectroscopy" or "colorimetry" or "immunofluorescence".	many	0	no

		If the variableMeasured is "depression rating", the measurementTechnique could be "Zung Scale" or "HAM-D" or "Beck Depression Inventory". If there are several variableMeasured properties recorded for some given data object, use a PropertyValue for each variableMeasured and attach the corresponding measurementTechnique. Bioschemas: To describe the process used to obtain a biological entity or which is associated with that entity (i.e procedure to obtain it or measure/characterise it)			
--	--	--	--	--	--

Based on schema.org [Thing]

Property	Expected Type	Description	CN	MG	CV
additionalProperty	PropertyValue	A property-value pair representing an additional characteristics of the entity, e.g. a product feature or another characteristic for which there is no matching property in schema.org.	many	O	no
alternateName	Text	An alias for the item.	many	R	no
description	Text	A description of the item.	many	R	no
identifier	PropertyValue or Text or URL	The identifier property represents any kind of identifier for any kind of Thing, such as ISBNs, GTIN codes, UUIDs etc. Schema.org provides dedicated properties for representing many of these, either as textual strings or as URL (URI) links. See background notes for more details. Recommendation: identifiers.org whenever possible	one	M	no
image	ImageObject or URL	An image of the item. This can be a URL or a fully described ImageObject.	many	O	no
name	Text	The name of the item.	one	R	no
sameAs	URL	URL of a reference Web page that unambiguously indicates the item's identity. E.g. the URL of the item's Wikipedia page, Wikidata entry, or official website.	many	O	no
url	URL	URL of the item.	one	O	no

New properties for [Thing]

Property	Expected Type	Description	CN	MG	CV
isMentionedIn	Thing	CreativeWork, Dataset, collection mentioning this entity	one	R	no

		Inverse of:mentions			
--	--	---------------------	--	--	--

New properties for [BiologicalEntity]

Property	Expected Type	Description	CN	MG	CV
associatedDisease	MedicalCondition OR URL	Disease associated to this protein feature	many	O	no
biocoordinates	QuantitativeValue	Coordinates in a 1 or 2D space, for instance length/coordinates in a sequence. Usage (example): Use a QuantitativeValue with properties minValue and maxValue for a region or value for a site in a Protein. Use it only with value for a protein length.	many	R	no
biologicalType	Text	List with types preferably supported by BioSchemas-> enumeration list of values maintained on a wikipage (cf accessibilityAPI in schema.org for implementation). If the value is not on the list then the data will still be parsed but only generic properties will be validated. {population,individual,tissue,cell,molecule,protein,nucleic acid} For the protein case the only biological type in use is "protein"	many	M	no
crossReference	Thing	A pointer to another, somehow related entity. Usage: Whenever isBasedOn/isBasisFor, isPartOf/hasPart, citation or any other more specific does not work.	many	O	no
phenotype	Text, Url, PropertyValue	To associate a biological entity to phenotypic information , whether the entity presents the phenotype or causes it.	many	O	no
representation	Text or URL or PropertyValue	Representation of this entity. For instance, chemical structure or sequence	many	O	no
sample	BiologicalEntity or URL	Clarify usage...	many	O	no
taxon	URL	A url pointing to NCBI Taxonomy or a taxonomic resource	one	O	yes

Legend:

CN: Cardinality (one, many)

MG: Marginality (M: minimum; R: recommended; O: optional)

CV: Suggested controlled vocabularies (yes, no)

Examples

Schema.org [suggests](#) implementing metadata using JSON-LD, RDFa or Microdata. JSON-LD is the recommended format by Google, but any of these formats can be used for embedding information about tools in a web page or other online resource.

Example 1 - [JSON-LD] [Example description: UniProt protein]

See full example for [P00519](#)

Example 2 - [JSON-LD] [Example description: Sample]

<TODO>

Example 3 - [JSON-LD] [Example description: Phenotype]

```
<script type="application/ld+json">
{
  "@context": "http://schema.org",
  "@type": "BiologicalEntity",
  "biologicalType": "phenotype",
  "Name": "0000001",
  "identifier": "https://bip.earlham.ac.uk/data_tables?model=trait_scores&query[plant_scoring_units.id]=89583",
  "url": "https://bip.earlham.ac.uk/data_tables?model=trait_scores&query[plant_scoring_units.id]=89583",
  "isPartOf": {
    "@type": "Thing",
    "url": "https://bip.earlham.ac.uk/data_tables?model=plant_trials&query[id]=47"
  },
  "phenotype": {
    "name": "Leaf Caesium concentration; method: 3-5 leaves per plant were pooled and freeze-dried. These were then ground and digested in nitric acid in a microwave digester prior to analysis by inductively coupled plasma-mass spectrometry; scale: mg/kg",
    "dateCreated": "2016-06-06 16:07:54",
    "dateModified": "2016-06-14 16:19:59",
    "distribution": "https://bip.earlham.ac.uk/trial_scorings/47.zip",
    "location": "University Nottingham",
  }
}
```

```
"measurementTechnique": "The scoring system was developed to take into account the host plant response and the growth of the pathogen as measured by the amount and type of sporulation. The combination of these characters gave six interaction phenotype classes of host/pathogen in"
```

Example 4 - [JSON-LD] [Example description: InterPro protein annotation, e.g., domain]

<TODO>

Example 5 - [JSON-LD] [Example description: PDBe protein structure]

See full example for [4wa9](#) structure

Controlled Vocabularies

This section contains a list of fields that require a controlled vocabulary or enumeration and suggests what is acceptable for each.

[biologicalType]

Enumeration: "protein", "protein annotation", "protein structure", "gene", "phenotype", "sample", "[initiator methionine](#)", "[domain](#)", "[protein region](#)", "[motif](#)", "[metal binding](#)", "[site](#)", "[active site](#)", "conserved site", "posttranslational modification", "homologous superfamily", "[calcium binding](#)", "[binding site](#)", "[zinc finger](#)", "[DNA binding](#)", "[signal peptide](#)", "[transit peptide](#)", "[propeptide](#)", "[chain](#)", "[peptide](#)", "[modified residue](#)", "[glycosylation](#)", "[lipidation](#)", "[disulfide bond](#)", "[cross-link](#)", "[compositional bias](#)", "[coiled coil](#)", "[sequence conflict](#)", "[alternative sequence](#)", "[alpha helix](#)", "[beta strand](#)", "[turn](#)", "[topological domain](#)", "[transmembrane](#)", "[intramembrane](#)", "[mutagenesis](#)", "[repeat](#)", "unique peptide", "non-unique peptide", "[natural variant](#)", "[non-standard residue](#)", "[sequence uncertainty](#)", "[non-adjacent residue](#)", "[non-terminal residue](#)", "variety", "plant line"

[taxon]

Ontology: Any well-known taxonomic resource. NCBI Taxon or UniProt Taxon are the recommended ones.