

IMAGE metadata - sample specification

This document describes the specification for all sample metadata. The [experiment](#) document is also available.

In the sample context, we consider donor animals and tissue samples at the moment. All samples must be registered in BioSamples at EMBL-EBI as this sample archive has the best support for ‘child of’ and ‘derived from’ sample relationships. The NCBI BioSample database is a peer of the EMBL-EBI BioSamples, and they exchange data regularly. IMAGE samples should be registered in the EMBL-EBI BioSamples prior to data submission. This document describes the attributes which must be associated with any BioSamples submission as well as optional fields.

Sample metadata requirements

Most requirements are laid out like this:

- **attribute name** (*data type*) a brief description

The details of data types can be found [here](#).

BioSamples takes sample records with a set of attributes. Each attribute has a name and a value. It can also have ‘Units’, or a ‘Term Source’ and a ‘Term Source ID’. The Term Source and ID allow us to refer to entries in other databases or ontologies. This is fully described on the [BioSamples help pages](#).

Common

These attributes should be present on every sample record.

Mandatory:

- **Data source ID** (*text*) the id used in the original data source to identify the record
- **Alternative id** (*limited value*) the other identifier used in the original data source, e.g. accession, secondary id etc.
- **Project** (*text*) All samples are stated to be part of the IMAGE project, while still belonging to each individual gene bank. This tag is required to identify them in public databases Allowed values are:
 - IMAGE
- **Submission title** (*text*) A descriptive title for the submission
- **Material** (*ontology_id*) type of the sample, either animal(organism) or sample (specimen from organism) Allowed values are:
 - [Organism](#)
 - [Specimen from organism](#)

- **Person last name** (*text*) the last name of the submitter or contact for these samples, multiple persons can be provided
- **Person email** (*uri_value*) email of contact person
- **Person affiliation** (*text*) affiliation of person
- **Person role** (*text*) the role (type of the contribution) made by the person
- **Organization name** (*text*) Organization name, multiple can be supplied, but at least one must be the name of the gene bank
- **Organization role** (*ontology_id*) the role of the organization, e.g. biobank, funder
- **Gene bank name** (*text*) The name of the gene bank that the data of this submission belongs to, ideally use the value stored in EUGENA
- **Gene bank country** (*limited value*) Country in which the gene bank resides, particularly important when different countries use the same database infrastructure, e.g. CryoWeb. The controlled vocabulary for countries is from INSDC <http://www.insdc.org/documents/country-qualifier-vocabulary>
- **Data source type** (*limited value*) the type of the data source the current Inject tool can handle Allowed values are:
 - CryoWeb
 - CRB-Anim
 - Template
- **Data source version** (*text*) the version of the data source, suggested to be either the dump date or the version number used by the source
- **Species** (*ontology_id*) the species of the sample record

Recommended:

- **Submission description** (*text*) A brief description of the submission
- **Person first name** (*text*) the first name of the submitter or contact for these samples, multiple persons can be provided
- **Organization address** (*text*) The address of the organisation, multiple can be supplied
- **Organization country** (*text*) Country in which organization resides, the controlled vocabulary for countries is from INSDC <http://www.insdc.org/documents/country-qualifier-vocabulary> Allowed values are:
 - the controlled vocabulary for countries is from INSDC <http://www.insdc.org/documents/country-qualifier-vocabulary>

Optional:

- **Description** (*text*) A brief description of the sample
- **Person initial** (*text*) the initial of the submitter or contact for these samples, multiple persons can be provided
- **Organization uri** (*uri_value*) the websites of organization, multiple can be supplied
- **Publication DOI** (*doi*) Publication associated with sample, supplied as the digital object identifier which can unambiguously point to the publication

Animal

An animal sampled for IMAGE. The following attributes are in addition to the attributes listed in the ‘Common’ section above.

Mandatory:

- **Supplied breed** (*text*) the breed information contained in the original data source, could be a local language or breed variation
- **EFABIS Breed country** (*text*) the country where the breed locates, the country name should be in the INSDC list
- **Sex** (*ontology_id*) the sex of the animal
- **Birth location accuracy** (*limited value*) Allowed values are:
 - missing geographic information
 - country level
 - region level
 - subregion level
 - precise coordinates

Recommended:

- **Mapped breed** (*ontology_id*) If possible breed will be mapped to a standardised breed ontology, both supplied and mapped breed will be available to end users
- **Birth location** (*text*) name of the place where the animal was given birth
- **Birth location longitude** (*number*) the longitude of the place where the animal was given birth in decimal degrees
- **Birth location latitude** (*number*) the latitude of the place where the animal was given birth in decimal degrees

Optional:

- **Child of** (*BioSample ID or Alternative animal ID*) Father and/or Mother’s animal ID whose record must also be present as part of the submission or BioSamples ID if sample has been previously submitted

Sample

A piece of tissue taken from an animal. The following attributes are in addition to the attributes listed in the ‘Common’ section above.

Mandatory:

- **Derived from** (*text*) The data source id or alternative id of the animal from which the sample was collected
- **Collection date** (*date*) Date when the sample was collected
- **Collection place** (*text*) name of the place where the sample was collected
- **Collection place accuracy** (*limited value*) Allowed values are:
 - missing geographic information

- country level
- region level
- subregion level
- precise coordinates
- **Organism part** (*ontology_id*) Ontology for organism part where the specimen was taken from

Recommended:

- **Specimen collection protocol** (*uri*) A URL link to the protocol used to collect this sample
- **Collection place latitude** (*number*) latitude of the place where the sample was collected
- **Collection place longitude** (*number*) longitude of the place where the sample was collected
- **Developmental stage** (*ontology_id*) developmental stage of the animal from which the specimen is collected
- **Physiological stage** (*ontology_id*) physiological stage of the animal from which the specimen is collected
- **Availability** (*uri_value or text*) Either an e-mail address to contact about availability or a link to a web page giving information on who to contact. E-mail addresses should be prefixed with ‘mailto:’, e.g. ‘<mailto:samples@example.ac.uk>’. In either case, long term support of the web page or e-mail address is necessary. Group e-mail addresses are preferable to individual accounts. Allowed values are:
 - no longer available
 - <mailto:a.valid@email>
- **Sample storage** (*text*) How the sample was stored. Temperatures are in degree Celsius. ‘Frozen, vapor phase’ refers to storing samples above liquid nitrogen in the vapor. Allowed values are:
 - ambient temperature
 - cut slide
 - frozen, -80 degrees Celsius freezer
 - frozen, -20 degrees Celsius freezer
 - frozen, liquid nitrogen
 - frozen, vapor phase
 - paraffin block
 - RNAlater, frozen -20 degrees Celsius
 - TRIzol, frozen
 - paraffin block at ambient temperatures (+15 to +30 degrees Celsius)
 - freeze dried
- **Sample storage processing** (*text*) How the sample was prepared for storage. Allowed values are:
 - cryopreservation in liquid nitrogen (dead tissue)
 - cryopreservation in dry ice (dead tissue)
 - cryopreservation of live cells in liquid nitrogen
 - cryopreservation, other

- formalin fixed, unbuffered
- formalin fixed, buffered
- formalin fixed and paraffin embedded
- freeze dried (variable for reproduction)
- freeze dried

Optional:

- **Animal age at collection** (*number*) Animal age at the point of collection.
- **Sampling to preparation interval** (*number*) How long between the sample being taken and the assay experiment preparations commencing. If sample preparations were then left in intermediate stages after preparation commenced, for example as sheared chromatin, then this should be made clear in your protocols