Data types for FAANG attributes

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. The following section describes the expectations for each data type within FAANG.

date

Dates should be reported in an ISO 8601 format, YYYY-MM-DD for dates or YYYY-MM for months. To ensure clarity, the format must be reported as the 'units'.

NCBI taxon ID

A species name and identifier from the NCBI Taxonomy database. For example, human would be described in the term with value of 'Homo sapiens', term source as 'NCBI Taxonomy' and term source ID as 9606.

number

A number, with units specified. BioSamples recommends that units are given without abbreviations. Terms defined in the UO are encouraged to be used. For example, a birth weight could have a value of 1.3 and the units specified as 'kilogram' .

protocol

A URL link to a protocol document on the FAANG FTP site. Please contact the FAANG data coordination centre to have your protocol documents added to the FTP site.

text

Text, using US English spellings.

URL

A URL, such as 'http://faang.org/'. Depending on the context, http, ftp, mailto links may be appropriate. Examples:

- ftp, ftp://ftp.faang.ebi.ac.uk/ftp/README
- http, http://faang.org/
- mailto, mailto:bob@example.org

ontology term

A reference to an ontology term. The attribute value should be the term label. The term source should be the ontology used, and the term source ID should be an ID from that ontology. For example, cerebral cortex could be described with an ontology term from 'UBERON' with ontology ID of 'UBERON:0000956' and the attribute value is 'cerebral cortex'. Though in the experiment submission, direct links to ontologies cannot be submitted as attributes. The use of ontology terms is still encouraged by setting the attribute value to exactly match the term name in the ontology.

location

A location should be reported as using three attributes:

- `location` (*text*) name of the location
- `location latitude` (*number*) latitude in decimal degrees. Units should be reported as 'decimal degrees'
- `location longitude`(*number*) longitude in decimal degrees. Units should be reported as 'decimal degrees'

sample

Samples can be referred to in two ways. If the sample you need to reference is in the submission, use the sample name. If the sample was already submitted, use the BioSample ID (e.g. SAMEA2821491).