

FAANG metadata - overview

This document describes the principles and structure for the FAANG metadata guidance.

The main goal of the [FAANG](#) standards is to ensure all FAANG experiments, samples and analyses are well described and that description is well structured. This will ensure as many potentially confounding factors as possible are recorded. We support the [MIAME](#) and [MINSEQE](#) guidelines, and aim to convert them to a concrete specification.

We divide the metadata into 3 related categories

- [Samples](#)
- [Experiments](#)
- [Analyses](#)

The detailed specification for each category is presented in additional documents linked to above. Each experiment will reference at least one sample. Each analysis will reference one or more experiments.

Metadata should be represented atomically. Each piece of information should be in a separate record, with a clear label.

Where multiple records are related, data should not be duplicated between them, e.g. tissue sample records from the same animal should not duplicate the animal information, they should point to a record for that animal.

Across all categories, descriptive factors should use ontology terms. Our preferred ontologies are EFO and the ontologies it imports, although we will use others where necessary. Where appropriate terms are not available in the ontology, we will work with that ontology to have the term added, please contact faang-dcc@ebi.ac.uk for help with this.

e.g., tissue specimens can be described using terms from UBERON such as [lung](#), [UBERON:0002048](#)

Protocols will be stored separately from the metadata, with a standardised name and stable URL, see the [FAANG wiki](#) for guidance on protocol naming. The URL/name can be referenced from the metadata submission files, each protocol receives a timestamp in its name (the date when FAANG Data Coordination Centre receives the protocol) so that subsequent updates to the protocols can be distinguished. e.g. The November 2016 version of the Roslin protocol for isolation of RNA from frozen tissue would get a name like this:

`ROSLIN_SOP_Isolation_of_RNA_from_frozen_tissue_20161108`

Protocols may specify data to record during the experiment. Some of this should be stored as metadata. Where the data is numeric, the unit should be specified.

e.g. ChIP sonication fragmentation size range (bp), Antibody batch number.

It is best these are hosted in a location which will be available in the long term so locations such as lab pages are inadvisable as web addresses change and hosting goes away. FAANG is happy to host protocols in the [FAANG FTP site protocols directory](#), see the [FAANG wiki](#) for further guidance.

If you wish FAANG to host your protocols, please send pdf copies of your protocols to faang-dcc@ebi.ac.uk.

Please name your files using this convention INSTITUTE_SOP_PROTOCOLNAME_YYYYMMDD e.g DEDJTR_SOP_Cryofreezing_Tissue_20160317.pdf. This is a protocol for Cryofreezing tissue, the protocol was written on the 17th March 2016 and the protocol comes from an Australian institute connected to the Victoria Department of Economic Development, Jobs, Transport and Resources. If you have any questions about protocols and the form they should take, please email the [FAANG Animal, Samples and Assays group](#).

Archive Submission

The different archives we recommend can support all of our metadata at a basic level. Where ever possible we recommend archives which explicitly support the use of ontology records. If it isn't possible, our guidelines will explain how to add ontology information using key value pairs as supported by all our recommended archives.

| Data Type | Archive | Host Institution | Native ontology support |
|--------------------------------|--------------|------------------|-------------------------|
| Sample | BioSamples | EBI | Yes |
| Experiment - WGS/Exomes | ENA | EBI | No |
| Experiment - WGS/Exomes | SRA | NCBI | No |
| Experiment - Epigenomics | ArrayExpress | EBI | No |
| Experiment - Epigenomics | Geo | NCBI | ? |
| Experiment - Transcriptomics | ArrayExpress | EBI | No |
| Experiment - Transcriptomics | Geo | NCBI | ? |
| Analysis - Alignment | ENA | EBI | No |
| Analysis - Variant Calls | EVA | EBI | No |
| Analysis - signal/region calls | ? | ? | ? |
| Unstructured data | BioStudies | EBI | ? |

The appropriate metadata recommendations are covered in their data type documents as linked to at the top of this document.

Sample records should be submitted to BioSamples@EBI prior to the submission of experimental results. These sample records will be mirrored by BioSample@NCBI. The sample records should be referenced when submitting experimental results to the appropriate assay archive.

Further guidance can be found on the [FAANG wiki pages](#).