FAANG

Establishing Metadata Standards, Validation and the FAANG Data Portal

Peter Harrison
Data Coordination and Archiving Team
peter@ebi.ac.uk
http://data.faang.org
The role of the FAANG Data Coordination Centre

• Implement the metadata rules from the FAANG Metadata and Data Sharing committee.

• Provide tools to support depositors in meeting these standards and submitting data to public archives.

• Build and maintain the data portal: http://data.faang.org

• Provide data coordination support: faang-dcc@ebi.ac.uk

http://data.faang.org
Metadata standards

https://www.ebi.ac.uk/vg/faang/rule_sets/

- Provide rich Sample and Experiment metadata standards for the FAANG community.
- Terminology controlled through ontologies to make downstream search and analysis more powerful.
- Enhancing reproducibility, accelerating research and enabling cross-depositor analyses to be performed.

Rules 15 rules

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Type</th>
<th>Required?</th>
<th>Allow multiple?</th>
<th>Valid values</th>
<th>Valid units</th>
<th>Valid terms</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Organism</td>
<td>NCBI taxon ID of organism</td>
<td>ncbi_taxon</td>
<td>mandatory</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>NCBITaxon_1 (include descendants)</td>
</tr>
<tr>
<td>Sex</td>
<td>Animal sex, described using any child term of PATO_0000047</td>
<td>ontology ld</td>
<td>mandatory</td>
<td>No</td>
<td>PATO_0000047</td>
<td></td>
<td></td>
<td>(include descendants)</td>
</tr>
<tr>
<td>birth date</td>
<td>Birth date, in the format YYYY-MM-DD, or YYYY-MM where only the month is known. For embryo samples record 'not applicable'</td>
<td>date</td>
<td>recommended</td>
<td>No</td>
<td>YYYY-MM-DD, YYYY-MM, YYYY</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Metadata and Data Sharing Committee update

• Version 3.5: [https://github.com/FAANG/faang-metadata](https://github.com/FAANG/faang-metadata)

• Updated WGS and HiC standards.

• Released Legacy and pools of specimens data standards.

• Carl (M&DS co-chair) and I are always happy to discuss anything about FAANG metadata.

• FAANG members should comply with FAANG Data sharing commitment: [https://www.faang.org/data-share-principle](https://www.faang.org/data-share-principle)

• If you are struggling to meet the standards please contact us, we are here to help: faang-dcc@ebi.ac.uk
Validation and conversion tools

https://www.ebi.ac.uk/vg/faang/validate/

• Validation and conversion tooling to assist depositors in meeting FAANG metadata standards.

• Depositors complete spreadsheet templates, validate them and convert to a format ready for archive submission.

Metadata validation

Rule set: FAANG Samples
Metadata file: Choose file, No file chosen
File format: BioSample.xlsx
Output format: Web page

Validate

http://data.faang.org
FAANG and legacy metadata standard tagging

• All data in the FAANG portal is revalidated against metadata standards and tagged as having met the full ‘FAANG’ or ‘Legacy’ standard.

• Important that you check that your data is in the FAANG data portal and is marked as meeting the FAANG standard. Please contact faang-dcc@ebi.ac.uk if it’s not.

• We will soon be importing all publically available data that meets the FAANG legacy standards from the archives.
FAANG Data portal

http://data.faang.org

Data Portal

FAANG is the Functional Annotation of Animal Genomes project. We are working to understand the genotype to phenotype link in domesticated animals.

This data portal will help find and browse FAANG’s data. Let us know what you think at faang-dcc@ebi.ac.uk.

Using this site

The Organisms page lets you search for any organism in the FAANG data set. We use the term “organism” to mean any individual animal that has contributed a biological specimen. Click an item in the list to see full details of the organism.

The Specimens page lets you search for any specimen in the FAANG data set. We use the term “specimen” to mean any biological material originating from a particular organism. Click an item in the list to see full details of the specimen.

The File page lets you search for any experiment files in the FAANG data set. We use the term “file” to mean any experiment assay file that has been submitted to a public archive. Click an item in the list to see full details of the file.

The Search page lets you search for both organisms and specimens contained in the FAANG data set. It returns results based on the text you enter. Click an item in the results list to see full details.

Contribute to FAANG

Sign up to become part of the FAANG community and join the working committees.

Instructions on how to submit your samples and analyses to FAANG are detailed on the FAANG wiki site archive submission guidelines.

It's good to share

FAANG is committed to sharing data rapidly, before publication. All members have agreed to the FAANG Data Sharing Statement for pre-publication data release.

FAANG is also committed to making regular public releases of primary and secondary analysis results to provide access to the wider community.
Data portal

http://data.faang.org

- Access all FAANG data (also legacy data soon).
- Apply ontology filters to find data of interest.
- Export table to file, useful for mass file download.

http://data.faang.org
Data portal

http://data.faang.org

• Individual detail pages with all metadata.
• Links to external ontology databases.
• Access sample and experiment protocols.
• Download any experiment files associated with the organism or sample.
Data portal

Direct links to data files in archives

http://data.faang.org/help/api

http://data.faang.org
Data portal

http://data.faang.org/search

• Predictive search.
• Returns organisms, specimens and files.
• Searches all metadata fields.
• Clickable links to individual detail pages.

http://data.faang.org
EMBL-EBI Unified Submissions Interface

- Single submissions interface for 8 EMBL-EBI archives, same system for samples, experiments and analyses.
- Built in validation system, will have all FAANG rulesets available.
- Rolling out to cover archives over coming year, BioSamples for sample submissions will be first.
- Backwards compatible, graduated switch to the new system when you are ready.
- More robust system with user experience lead design.
- Extensive training, documentation and help will be available.

http://data.faang.org
FAANG data submission training workshops (proposed not confirmed)

1. 7th ISAFG, Adelaide, Australia – November 2018
2. PAG XXVII, San Diego, USA – January 2019
3. ISAG, Lleida, Spain – July 2019

• Validation against FAANG standards.
• Utilising new Unified Submissions Interface.
• Selecting appropriate ontologies for your data.
• Sample, experiment and analysis submissions.
• Interactive hands on training.
• Bring your own metadata!

http://data.faang.org
Planned developments

• Full integration with new EMBL-EBI Unified Submissions Interface.
• Trackhubs and ENSEMBL browser integration.
• Data portal user experience (UX) improvements.
• Archive legacy data available from the FAANG portal.
• Develop FAANG variation data and analysis data submission support tools.
• Provide display view window of FAANG studies/consortiums by group and organism.
• Programmatic access (API) improvements.

http://data.faang.org
We are here to help with your data...

http://data.faang.org

faang-dcc@ebi.ac.uk
peter@ebi.ac.uk

Come find me - EMBL-EBI Stand 405, Grand Exhibit Hall