

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

The screenshot shows the 'Metadata validation' tool. It includes a 'Validate' button and a 'Rule set' dropdown menu set to 'FAANG Samples'. Below the form, there is a link to 'The latest sample metadata Excel file templates and guidance on how to complete them can be found on the FAANG wiki site.' At the bottom, it mentions 'FAANG metadata tools' and 'Tools for validating FAANG metadata, developed and hosted by EMBL-EBI'.

The screenshot shows the 'FAANG organisms' page. It features a navigation bar with 'Home', 'Organisms', 'Specimens', 'Files', 'Search', and 'Help'. Below the navigation, there are 'Export as CSV file' and 'Export as Tabular file' buttons. The main content is a table with columns for 'Standard', 'BioSample ID', 'Sex', 'Organism', and 'Breed'. The table lists various samples with their corresponding IDs, sexes, organisms, and breeds.

Standard	BioSample ID	Sex	Organism	Breed
FAANG 284	SAMEA799918	female	Bos taurus	Holstein
FAANG Legacy 33	SAMEA7999168	female	Sus scrofa	Large White
Sex +	SAMEA7997668	female	Bos taurus	Holstein
male 205	SAMEA7996168	female	Capra hircus	Alpine
female 100	SAMEA7993918	female	Capra hircus	Alpine
Organism +	SAMEA7992418	male	Bos taurus	Holstein
Gallus gallus 132	SAMEA7989418	male	Bos taurus	Holstein
	SAMEA7988668	male	Capra hircus	Alpine

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								
Name	Description	Type	Required?	Allow multiple?	Valid values	Valid units	Valid terms	Condition
Organism	NCBI taxon ID of organism	ncbi_taxon	mandatory	No			NCBITaxon_1 (include descendants)	
Sex	Animal sex, described using any child term of PATO_0000047	ontology_id	mandatory	No			PATO_0000047 (include descendants)	
birth date	Birth date, in the format YYYY-MM-DD, or YYYY-MM where only the month is known. For embryo samples record 'not applicable'	date	recommended	No		'YYYY-MM-DD' 'YYYY-MM' 'YYYY'		

Metadata and Data Sharing Committee update

- Version 3.5: <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>
- If you are struggling to meet the standards please contact us, we are here to help: faang-dcc@ebi.ac.uk



Carl Schmidt
M&DS Co-chair

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



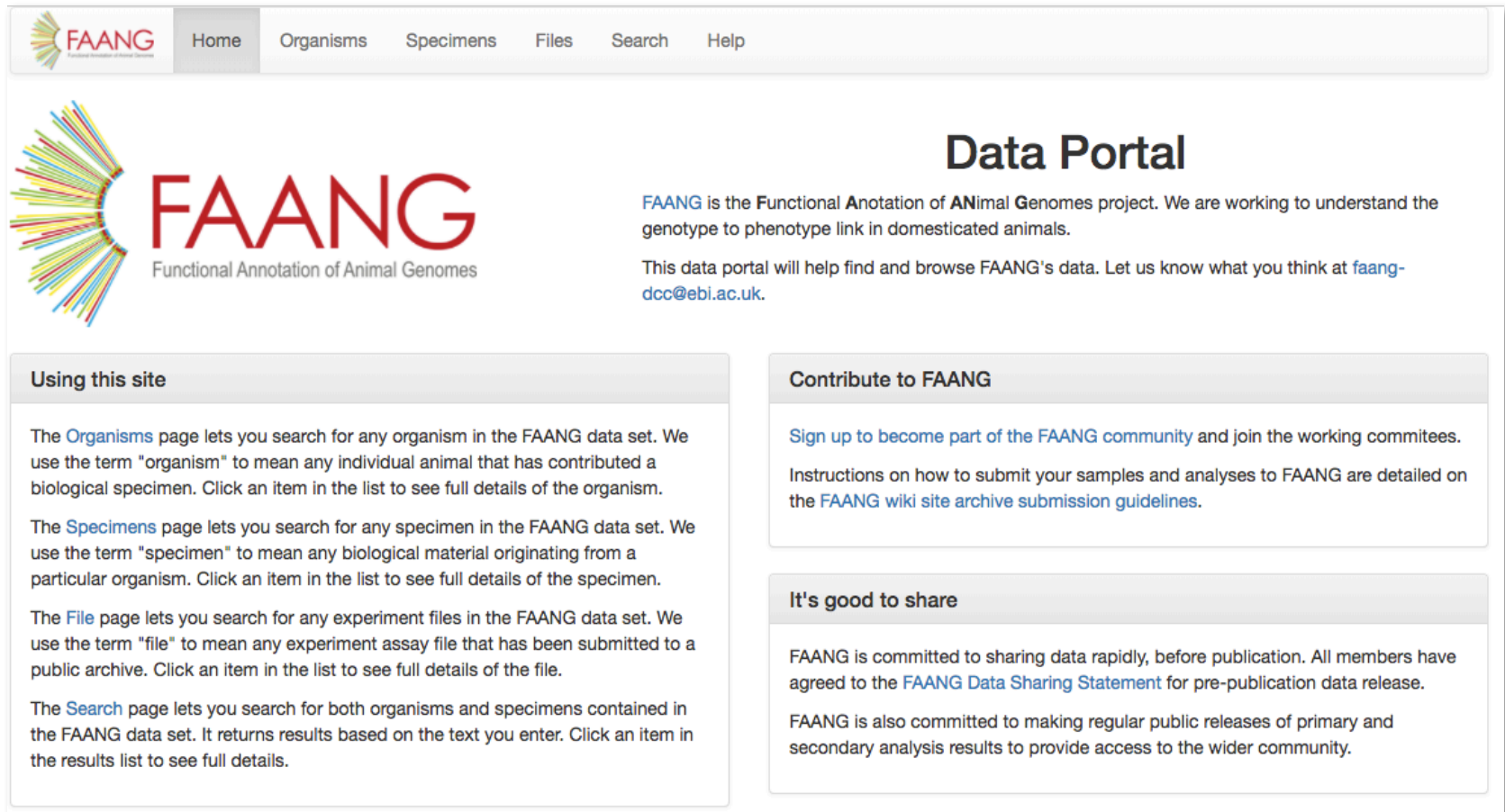
BTA_NMBU_S36	pass	sample	• standard	Monocyte-derived macrophages stimulated with Streptococcus agalaciae strain ST103	blood specimen
M1	error	sample	• standard • purified cells	Monocyte-derived macrophages unstimulated	cell specimen

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 its not.
- We will soon be importing all publically available data that meets the FAANG legacy standards from the archives.



<http://data.faang.org>



The screenshot shows the FAANG Data Portal website. At the top is a navigation bar with links for Home, Organisms, Specimens, Files, Search, and Help. The main content area features the FAANG logo and the text: "Data Portal" and "FAANG is the Functional Annotation of ANimal Genomes project. We are working to understand the genotype to phenotype link in domesticated animals." Below this, it says "This data portal will help find and browse FAANG's data. Let us know what you think at faang-dcc@ebi.ac.uk." There are three main sections: "Using this site" which describes the Organisms, Specimens, File, and Search pages; "Contribute to FAANG" which encourages signing up and provides submission guidelines; and "It's good to share" which states FAANG's commitment to data sharing and public releases.

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

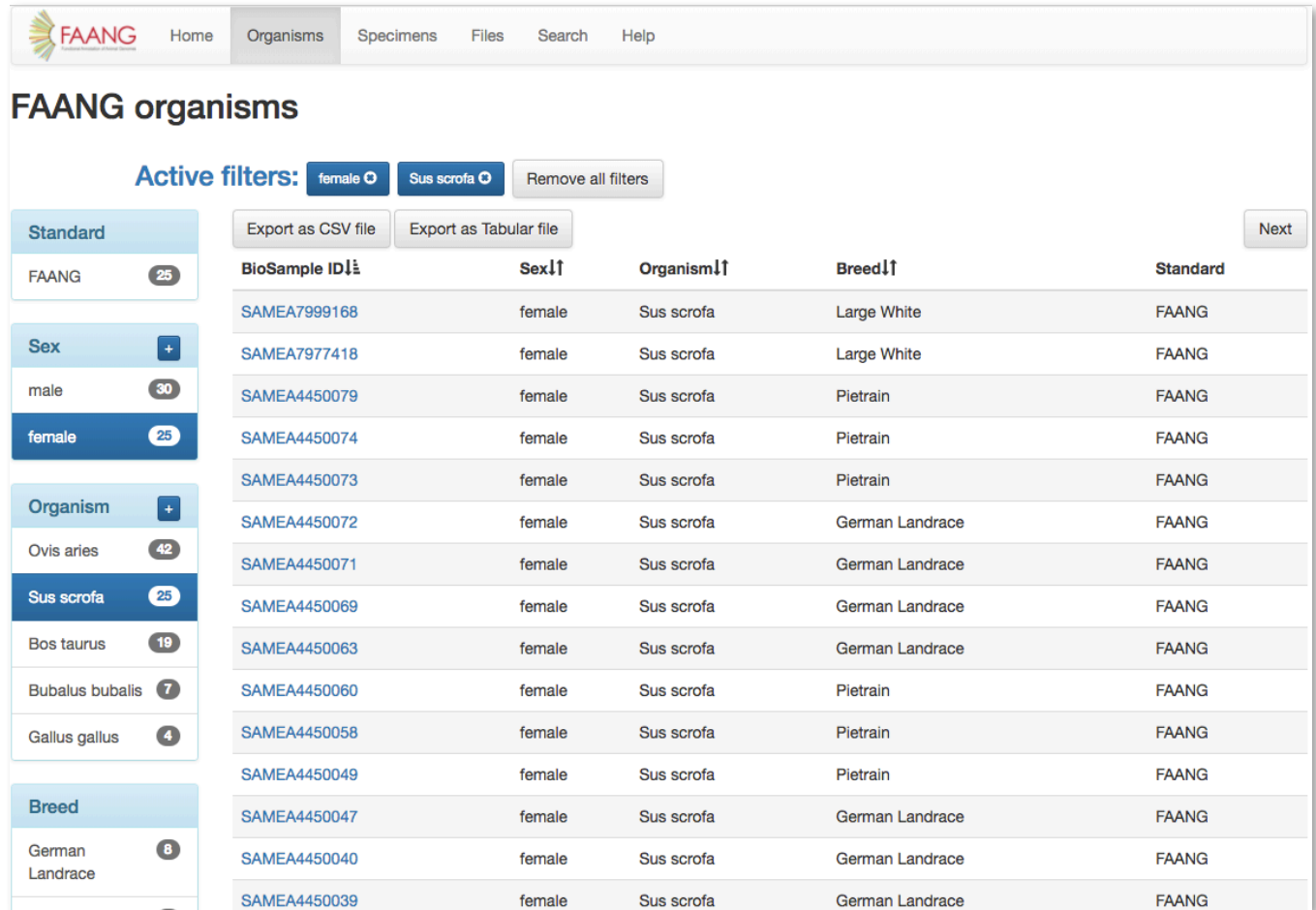
FAANG is committed to sharing data rapidly, before publication. All members have agreed to the [FAANG Data Sharing Statement](#) for pre-publication data release.

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.



FAANG Home Organisms Specimens Files Search Help

FAANG organisms


Active filters: female 25 Sus scrofa 25 Remove all filters

Export as CSV file Export as Tabular file Next

BioSample ID	Sex	Organism	Breed	Standard
SAMEA7999168	female	Sus scrofa	Large White	FAANG
SAMEA7977418	female	Sus scrofa	Large White	FAANG
SAMEA4450079	female	Sus scrofa	Pietrain	FAANG
SAMEA4450074	female	Sus scrofa	Pietrain	FAANG
SAMEA4450073	female	Sus scrofa	Pietrain	FAANG
SAMEA4450072	female	Sus scrofa	German Landrace	FAANG
SAMEA4450071	female	Sus scrofa	German Landrace	FAANG
SAMEA4450069	female	Sus scrofa	German Landrace	FAANG
SAMEA4450063	female	Sus scrofa	German Landrace	FAANG
SAMEA4450060	female	Sus scrofa	Pietrain	FAANG
SAMEA4450058	female	Sus scrofa	Pietrain	FAANG
SAMEA4450049	female	Sus scrofa	Pietrain	FAANG
SAMEA4450047	female	Sus scrofa	German Landrace	FAANG
SAMEA4450040	female	Sus scrofa	German Landrace	FAANG
SAMEA4450039	female	Sus scrofa	German Landrace	FAANG

Standard: 25
Sex: male 30, female 25
Organism: Ovis aries 42, Sus scrofa 25, Bos taurus 19, Bubalus bubalis 7, Gallus gallus 4
Breed: German Landrace 8

SAMEA103886174

Name: SUS_RI_DUR21-09
BioSample ID: [SAMEA103886174](#)
Release date: 2016-10-14
Update date: 2016-10-25
Sex: female
Organism BioSample Id: [SAMEA103886165](#)
Organism: [Sus scrofa](#)
Breed: [Duroc](#)
Health status: [normal](#)
Description: kidney cortex
Standard met: FAANG
Project: FAANG
Organisation: [The Roslin Institute and Royal Dick School of Veterinary Studies](#) (biomaterial provider) [BBSRC](#) (funder) [EMBL-EBI](#) (curator) [The Roslin Institute and Royal Dick School of Veterinary Studies](#) (institution)
Material: [specimen from organism](#)
Derived from: [SAMEA103886165](#)
Specimen collection date: 2013-08-01
Animal age at collection: 4 month
Developmental stage: [juvenile stage](#)
Organism part: [cortex of kidney](#)
Specimen collection protocol: [ROSLIN_SOP_Harvest_of_Large_Animal_Tissues_20160516.pdf](#) 
Fasted status: fed

Files

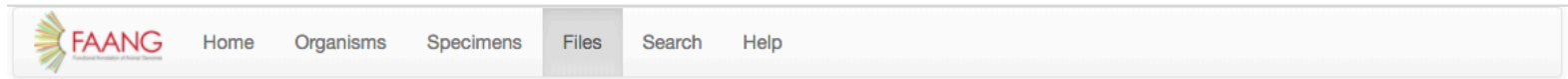
Next

Name	Archive	Experiment	Run	Download
ERR1824263_1	ENA	ERX1886667	ERR1824263	
ERR1824263_2	ENA	ERX1886667	ERR1824263	

<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.

Direct links to data files in archives



FAANG files

Study	File name	Study	Experiment	Species	Assay type	Specimen	Instrument
PRJEB19386 610	ERR1017174_1	PRJEB9561	ERX1096241	Sus scrofa		SAMEA3540911	Illumina HiSeq 2000
PRJEB14330 126	ERR1017174_2	PRJEB9561	ERX1096241	Sus scrofa		SAMEA3540911	Illumina HiSeq 2000
PRJEB8784 99	ERR1017177_1	PRJEB9561	ERX1096244	Sus scrofa		SAMEA3540912	Illumina HiSeq 2000
PRJEB9561 20	ERR1017177_2	PRJEB9561	ERX1096244	Sus scrofa		SAMEA3540912	Illumina HiSeq 2000
	ERR1017178_1	PRJEB9561	ERX1096245	Sus scrofa		SAMEA3540912	Illumina HiSeq 2000
	ERR1017178_2	PRJEB9561	ERX1096245	Sus scrofa		SAMEA3540912	Illumina HiSeq 2000

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

How can I access data programmatically?

In addition to the data portal, FAANG also provides APIs which allow programmatic access to FAANG data.

Examples

Get all available animals:

```
curl "http://data.faang.org/api/organism/_search/?fields=_id&size=1000"
```

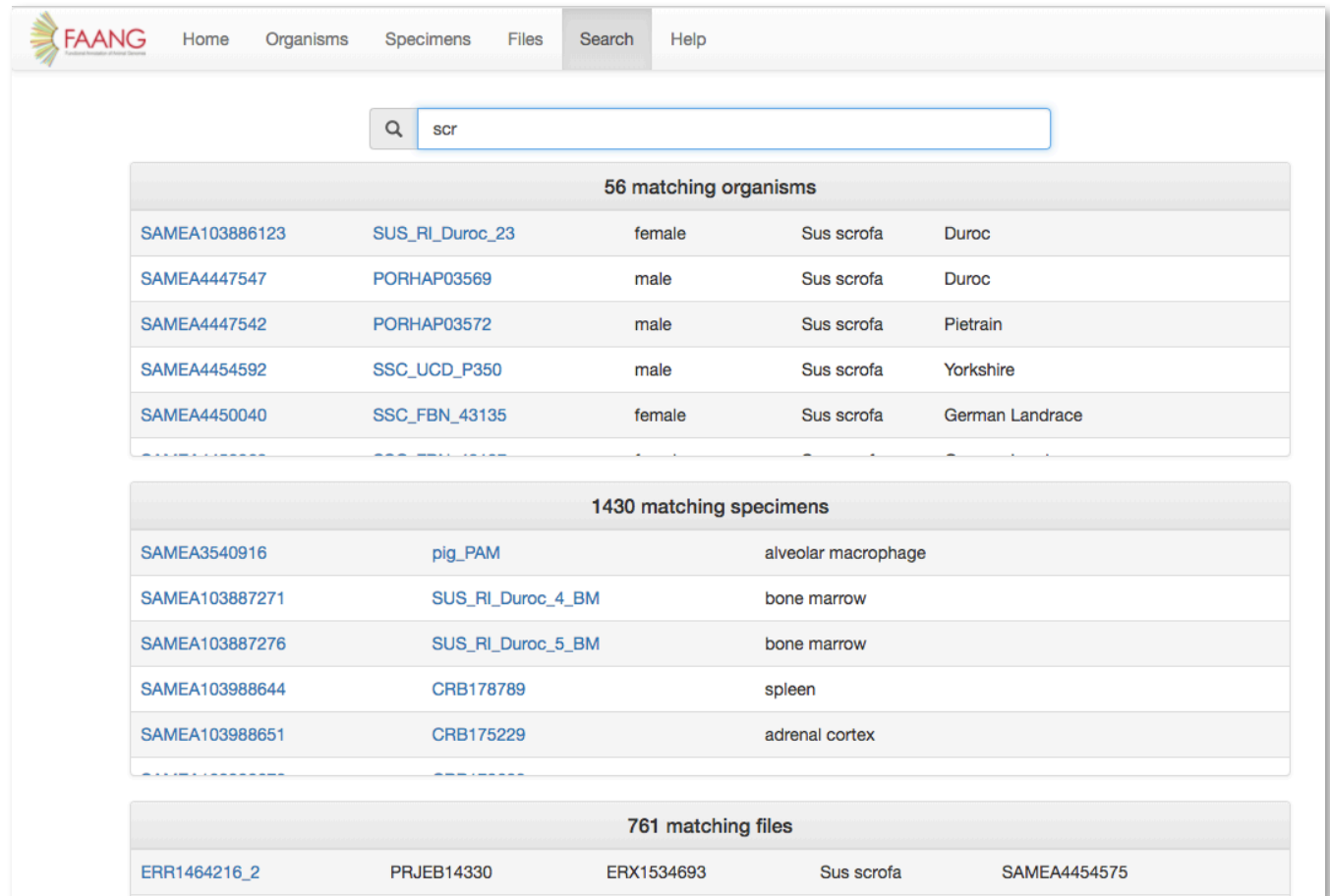
Get information for animal with BioSample accession SAMEA7999918:

```
curl "http://data.faang.org/api/organism/SAMEA7999918"
```

Data portal

<http://data.faang.org/search>

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



The screenshot shows the FAANG search interface with a search bar containing 'scr'. Below the search bar, there are three sections of results:

- 56 matching organisms**: A table with 5 rows showing organism details. The first row is: SAMEA103886123, SUS_RI_Duroc_23, female, Sus scrofa, Duroc.
- 1430 matching specimens**: A table with 5 rows showing specimen details. The first row is: SAMEA3540916, pig_PAM, alveolar macrophage.
- 761 matching files**: A table with 1 row showing file details: ERR1464216_2, PRJEB14330, ERX1534693, Sus scrofa, SAMEA4454575.

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.



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!



Planned developments

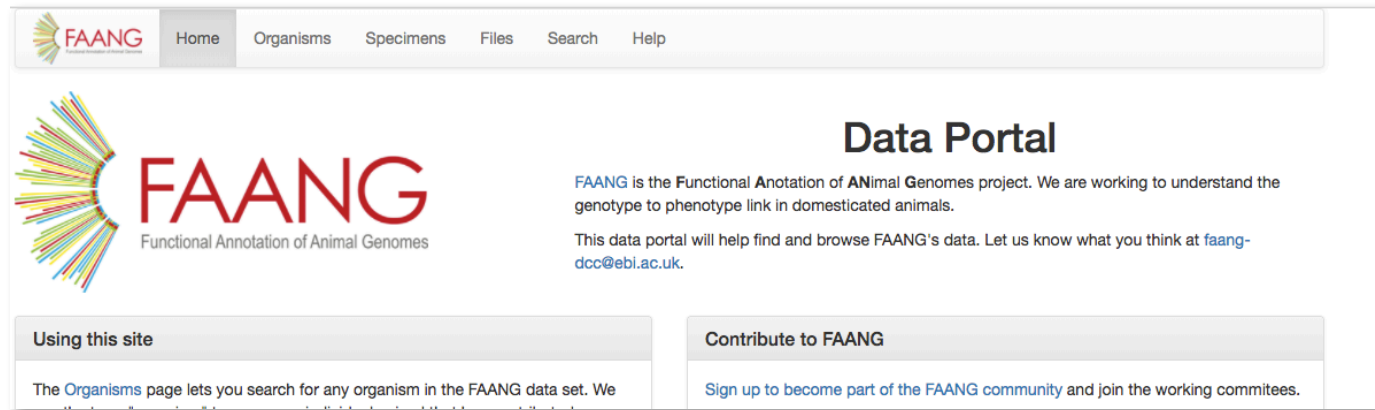
- Full integration with new EMBL-EBI Unified Submissions Interface.
- Trackhubs and ENSEMBL browser integration.
- Data portal user experience (UX) improvements.
- Archive legacy data available from the FAANG portal.
- Develop FAANG variation data and analysis data submission support tools.
- Provide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



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



<http://data.faang.org>

