



**Identification of functionally active genomic features relevant to phenotypic diversity and plasticity in cattle**

# Deliverable 10.1

## Data Management Plan

**Grant agreement no°: 815668**

Due submission date

**2020-02-29**

Actual submission date

**2020-02-28**

Responsible author(s)

**Peter Harrison** [peter@ebi.ac.uk](mailto:peter@ebi.ac.uk)

**Daniel Zerbino** [zerbino@ebi.ac.uk](mailto:zerbino@ebi.ac.uk)

**Confidential No**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 815668. The content of this report reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

**DOCUMENT CONTROL SHEET**

Deliverable name	Data Management Plan
Deliverable number	10.1
Partners providing input to this Deliverable	EMBL-EBI
Draft final version circulated by lead party to: On date	WP leader 2020-02-17
Approved by (on date)	FBN as Coordinator (2020-02-28)
Work package no	10
Dissemination level	Public (PU)

**REVISION HISTORY**

Version number	Version date	Document name	Lead partner
Vs 1	2020-02-14	D10.1_Data_Management_Plan	EMBL-EBI

**Changes with respect to the DoA (Description of Action)**

N/A

**Dissemination and uptake**

Public

## Table of Content

<b>SUMMARY OF RESULTS</b>	<b>3</b>
<b>INTRODUCTION</b>	<b>3</b>
<b>CORE REPORT</b>	<b>3</b>
<b>CONCLUSIONS</b>	<b>4</b>
<b>ANNEXES</b>	<b>4</b>

# Summary of results

The purpose of this deliverable was to prepare the first version of the Data Management Plan (DMP) for the BovReg project. The Data Management Plan will be updated periodically throughout the project and at least at the end of each reporting period. The BovReg DMP was written using the template and guidance provided by H2020, ensuring that all required information was covered in the text. The dataset sizes are not yet known so these will be included in an updated version of the DMP. The DMP is an important document for reference by the entire consortium, and the latest version will always be made available to the consortium to ensure the latest information on data management and FAIR data access is available to those generating and analysing BovReg data. The DMP was written by Peter Harrison and Daniel Zerbino (EMBL-EBI), with input from the wider consortium.

# Introduction

The first version of the BovReg DMP was written and distributed to the consortium. It provides guidance to the entire consortium on the appropriate management of BovReg data throughout the project and preparation for data persistence beyond the project end date.

# Core report

Version 1 of the BovReg DMP is attached in annex 5.1, and has also been distributed to all members of the consortium. The DMP was prepared using the template and guidance provided by H2020. The DMP covers the different release strategies of BovReg data, its commitment to releasing data openly through public archives and ensuring that the data meets FAIR data standards. As BovReg is a member of the larger FAANG consortium coordinated action, the DMP also covers the strategies that BovReg will employ to meet the requirements of FAANG in terms of metadata standards and data release. The DMP ensures that all BovReg data will be appropriately prepared and distributed to conform

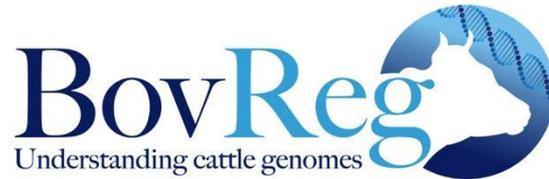
with FAIR (findable, accessible, interoperable and reusable) standards. The DMP outlines the strategies to ensure long term data security and preservation beyond the end of the project by utilising the separately funded EMBL-EBI public archives as the persistent open data store for all BovReg generated data.

## **Conclusions**

The first version of the BovReg DMP was approved and distributed to all members of the consortium. The DMP will be periodically updated throughout the project, with all subsequent updates also distributed to all consortium partners.

## **Annexes**

See next page.



**Project Number: 815668**

**Project Acronym: BovReg**

**Project title: BovReg - Identification of functionally active genomic features relevant to phenotypic diversity and plasticity in cattle**

## **DATA MANAGEMENT PLAN**

**Version: 1.0**

**Authors: Peter Harrison and Daniel Zerbino (EMBL-EBI)**

The information in this document is provided as is and no guarantee or warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

<b>1. Data Summary</b>	<b>2</b>
<b>2. FAIR data</b>	<b>6</b>
2.1. Making data findable, including provisions for metadata	6
2.2. Making data openly accessible	7
2.3. Making data interoperable	8
2.4. Increase data re-use (through clarifying licences)	10
<b>3. Allocation of resources</b>	<b>11</b>
<b>4. Data security</b>	<b>12</b>
Omic Data	12
Personal data	12
<b>5. Ethical aspects</b>	<b>13</b>
Omics data	13
Personal data	13
<b>6. Other issues</b>	<b>13</b>

## 1. Data Summary

The Data Management Plan will be periodically updated throughout the project to reflect changes in the data produced by the project and any changes in storage and release. A future update will include the sizes of the datasets produced by the project.

### 1.1. Omics Data

The Identification of functionally active genomic features relevant to phenotypic diversity and plasticity in cattle (BovReg) project aims to provide a comprehensive map of functionally active genomic features in cattle and how their (epi)genetic variation in beef and dairy breeds translates into phenotypes. The project will generate functional genome data based on the core recommended assays of the FAANG consortium (Andersson et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol* **16**, 57 (2015).

<https://doi.org/10.1186/s13059-015-0622-4>) and conform to the data recording standards set by the FAANG consortium (<https://data.faang.org/ruleset/samples#standard>). This will provide the FAANG community with a catalogue of functionally active genomic features relevant to phenotypic diversity and plasticity in cattle.

The BovReg consortium will handle three large categories of omics data: datasets generated as part of BovReg, background data contributed by the partners and analysis results derived from the two previous categories.

The datasets generated as part of the BovReg project fall under two data release strategies based on whether the data in question was generated and analysed by project partners that are part of the global FAANG consortium (<https://www.faang.org/contributors>) or not. Data produced or analysed by partners that are part of the global FAANG consortium will adhere to the FAANG Data Sharing Statement (<https://www.faang.org/data-share-principle>), and thus will be made available as soon as possible through the public archives and FAANG data portal. This data will be available pre-publication in accordance with the FAANG data reuse statement. Data produced by partners that are not part of the global FAANG consortium will be initially only available to consortium partners through an authenticated BovReg private data hub until the datasets are used in a publication or at the end of the project, whichever comes first. A list of which BovReg generated datasets fall under which data release strategy is listed below. Regardless of the data release strategy by the end of the project datasets newly generated omics within the BovReg project will be openly distributed in accordance with open FAIR data sharing standards.

Background datasets contributed by partners will be handled according to the conditions set by their producers. In some cases, this data may already be part of FAANG and therefore freely available to FAANG members. However, we will also be handling private datasets that were produced outside of FAANG. These mostly concern genotypic and phenotypic data for GWAS and QTL mapping that are not covered by FAANG. These data will be shared privately among BovReg partners for meta-analysis.

Derived analysis results produced under BovReg will be handled on a case-by-case basis to ensure that our commitments to the data providers, the Open Research Data Pilot ([https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)) and the FAANG community are simultaneously respected. In

particular, the meta-analysis of private genotype datasets will only be released as needed to validate results presented in scientific publications.

A. The following datasets are generated by FAANG consortium partners and will be shared pre-publication in accordance with the FAANG Data Sharing Statement:

- The sample metadata for CAGE, Hi-C, ChIRP, ATAC-Seq, Chip-Seq, WGBS, RRBS and RNAseq assays conducted within BovReg from all partners will be made available through the BioSample archive, the FAANG data portal (<https://data.faang.org/>) and the public BovReg FAANG data portal (<https://data.faang.org/projects/bovreg>).
- CAGE Transcription Start Sites: Analysis of TSSs from a subset of tissues, produced in WP2.
- Hi-C interaction: Analysis of interactions between regions of chromatin, produced in WP2.
- ChIRP: Analysis of interactions between chromatin and long non-coding RNA, produced in WP2.
- WGBS and RRBS: Measurement of methylation level along the genome for a given cell type, produced in WP5.
- RNA-Seq: Analysis of transcription in GxE experiments, produced in WP5, analysis of RNAseq of immortal cell lines produced in WP1.

B. The following datasets will initially only be available to members of the consortium and later made publicly available upon paper publication or at the end of the project, whichever comes first. This data is produced by partners that are not part of the FAANG consortium and therefore are not subject to the FAANG data sharing principles.

- RNA-Seq and small RNA-Seq: Analysis of transcription from a subset of tissues, produced in WP2.
- ChIP-Seq: Measurement of chromatin marks and transcription factor binding along the genome for a given cell type, produced in WP2

- ATAC-Seq: Measurement of chromatin openness along the genome for a given cell type, produced in WP2 and WP5
- C. To date, the partners have already committed to contributing the following private background datasets (more datasets may be brought in as needed):
- Whole-genome sequence (WGS) data with depth of about 10X or more are available for Holstein (GIGA: 760), Belgian Blue (GIGA: 200), Brown Swiss (ETH: 60) and Original Braunvieh (ETH: 50), used in WP2.
  - Crossbred Metabolomic profiles (FBN): 3 data sets of targeted metabolic profiles from plasma samples (n= 180, n=150, n=60) from ChaxHF crossbred animals sampled at experimental station, used in WP4.
  - Brown Swiss and Original Braunvieh GWAS (ETH): Results of sequence-based GWAS for somatic cell count and milk production in Brown Swiss and Original Braunvieh cattle, used in WP 4 and 6.
  - 3,200 existing RNA-Seq datasets available to the BovReg partners from previous national and international projects, used in WP 4
  - 144 ATAC-seq, RRBS and RNA-seq libraries spanning three cattle breeds (Holstein-Friesian, NDama and Nelore) and seven immune cell types.
  - Blood metabolomic data from 1,100 individuals available to the BovReg partners from previous national and international projects, used in WP4.
  - Holstein GWAS (WR): Results of sequence-based GWAS for several traits will be used within WP4.
  - Feed intake data from Global Dry Matter Intake (gDMI), used in WP7
- D. The project will make use of existing open access datasets from FAANG and the wider public archives:
- Crossbred RNA-Seq (FBN): Transcriptome profiles from 60 Cha x HF crossbred animals from 4 tissues (liver, jejunum, muscle, rumen) provided by partner FBN, and used to evaluate gene expression in WP 4 and 5.

- Publicly available WGS data for 105 Fleckvieh, Angus and Jersey animals from Run2 of the 1000 Bull Genomes Project, used in WP2.
- Iso-Seq datasets for Hereford (NCBI BioProject IDs PRJNA386670), and other Iso-Seq datasets that will be available from Holstein (e.g. PRJNA434299) and Jersey, used in WP2.
- Transcriptomic datasets for sheep (UEDIN) (e.g. PRJEB19199, PRJEB34864, PRJNA414087), used for comparative analysis in WP3.

## 1.2 Human Subjects Data

Bovreg will be collecting 2 human subject datasets:

- Semi-structured interviews with scientists and experts, performed in WP8 task T8.1. All data will be published in anonymized form.
- Outputs of the Democs games, performed in WP8 task T8.2. These will either be entered in electronic form directly into the Democs pages of the BovReg website, or handwritten on to paper sheets, which will subsequently be scanned into electronic form for archiving. Data sharing will be facilitated through the University of Edinburgh's DataShare repository.

## 2. FAIR data

### 2. 1. Making data findable, including provisions for metadata

The omics datasets, once available in the public archives and FAANG data portal in accordance with the data release strategy described in section 1, will be highly discoverable by the community. BovReg will utilise the rich and extensive FAANG metadata standards for the documentation and description of its datasets

(<https://data.faang.org/ruleset/samples#standard>). All BovReg omics datasets will fully meet these standards, as verified by their submission preparation using the FAANG data validation software (<https://data.faang.org/validation/samples>). Each organism, sample and dataset will be assigned a globally recognised unique identifier by the EMBL-EBI public archives. These accessions are globally recognised by the comparable archives at the National Center for

Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>) and DNA Databank of Japan (DDBJ; <https://www.ddbj.nig.ac.jp/index-e.html/>). These identifiers and the public archives provide interconnectivity making it easy for a user to navigate and locate all of the datasets generated on a given sample. BovReg will conform with the record naming conventions of the FAANG consortium guidelines ([https://dcc-documentation.readthedocs.io/en/latest/biosamples\\_template/](https://dcc-documentation.readthedocs.io/en/latest/biosamples_template/)). The FAANG data portal (<https://data.faang.org/>) and the BovReg data portal view (<https://data.faang.org/projects/bovreg>) utilises an ElasticSearch backend to ensure that ontology metadata fields are keyword searchable using its predictive simultaneous search across organisms, samples, reads and analyses (<https://data.faang.org/search>). The data portal table views provide filters that allow a user to explore the data based on species, technology, breeds, sex, material, organism part, cell type, assay type, archive, and sequencing instrument. All software generated by BovReg will be appropriately versioned.

## 2.2. Making data openly accessible

BovReg data release is based on five different strategies as outlined in section 1. All public BovReg omics datasets will be deposited in the EMBL-EBI public archives, using the submission validation and submission system provided by the FAANG Data Coordination Centre. For datasets released from FAANG consortium partners the datasets will include a reserved right of first publication stipulation set out in the FAANG Data Sharing statement (<https://www.faang.org/data-share-principle>). However, once this has been met, there will be no restrictions on use of the data and no data access committee is required for BovReg datasets. The FAANG data portal includes anonymous usage analytics, but no tracking of individual data use will be made. The omics datasets will include the following Data Sharing Statement in the dataset description field:

*"This study is part of the FAANG project, promoting rapid prepublication of data to support the research community. These data are released under Fort Lauderdale principles, as confirmed in the Toronto Statement (Toronto International Data Release Workshop. Birney et al. 2009. Pre-publication data sharing. Nature 461:168-170). Any use of this dataset must abide by the FAANG data sharing principles. Data producers reserve the right to make the first publication of a global analysis of this data. If you are unsure if you are allowed to publish on this dataset, please*

*contact the FAANG Consortium ([faang@iastate.edu](mailto:faang@iastate.edu)) to enquire. The full guidelines can be found at <http://www.faang.org/data-share-principle>."*

The EMB-EBI public archives, as a project partner, are fully aware of the incoming data from the BovReg project, and fully prepared for accepting this data. BovReg omics data will be available for download by the community either direct from the public archives or via the FAANG data portal as direct download links. The EMBL-EBI public archives support the following open source access options: web browser download, FTP, Aspera, Globus and API access methods.

Software generated by the BovReg consortium will be publicly available from the FAANG GitHub repository (<https://github.com/FAANG/>) pre-publication. The software will be permissively licensed (e.g. Apache 2.0 or MIT), will include complete documentation, Nextflow workflow management and will use Docker containers to ensure deployability. No specific tools are required to access the data from the data portals or the FAANG data portal, as they will use standard accepted file formats of the public archives.

As described in section 1.1, data newly generated within the BovReg project in subsections 1.1.A and 1.1.B will be openly accessible upon publication or at the end of the project, with some data available publicly earlier under the FAANG sharing statement.

### 2.3. Making data interoperable

BovReg ensures that the samples and datasets generated during the project are interoperable once in the public domain by utilising the FAANG metadata standards. The BovReg private data hub utilises the same metadata standards as the FAANG data portal, ensuring that samples and data are accurately recorded early in the project for rapid release upon publication. BovReg will use the latest version of the FAANG metadata standards (<https://data.faang.org/ruleset/samples#standard>), and ensure accuracy by using the FAANG validation software for preparation of data to ensure compliance (<https://data.faang.org/validation/samples>). The omics datasets generated by BovReg will use standard bioinformatic data file formats (i.e. FASTQ, FASTA, SAM/BAM, GFF/GTF, BED, and VCF) as recommended by the relevant public archives at EMBL-EBI. Overall the project contributes

to the global FAANG coordinated effort with the deliverance of a high-quality functional genome maps for cattle.

Pipelines will be orchestrated using open workflow managers such as Nextflow (developed by project partners at CRG) and open software applications utilised within the pipelines will be deposited as Bioconda recipes to ensure consistent reuse across the project and by downstream users. BovReg will develop and adopt pipeline development standards to ensure reusability of the software generated in the project by third parties.

Interoperability is also ensured through the consistent use of ontological terms as recommended by the FAANG Data Coordination Centre and FAANG Metadata and Data sharing working group. A key goal of the project is to further the use and applicability of ontologies within cattle livestock. The ontologies that will initially be utilised in the project will be:

OBI	<a href="https://www.ebi.ac.uk/ols/ontologies/obi">https://www.ebi.ac.uk/ols/ontologies/obi</a>
NCBI Taxonomy	<a href="https://www.ebi.ac.uk/ols/ontologies/ncbitaxon">https://www.ebi.ac.uk/ols/ontologies/ncbitaxon</a>
EFO	<a href="https://www.ebi.ac.uk/ols/ontologies/efo">https://www.ebi.ac.uk/ols/ontologies/efo</a>
LBO	<a href="https://www.ebi.ac.uk/ols/ontologies/lbo">https://www.ebi.ac.uk/ols/ontologies/lbo</a>
PATO	<a href="https://www.ebi.ac.uk/ols/ontologies/pato">https://www.ebi.ac.uk/ols/ontologies/pato</a>
VT	<a href="https://www.ebi.ac.uk/ols/ontologies/vt">https://www.ebi.ac.uk/ols/ontologies/vt</a>
ATOL	<a href="https://www.ebi.ac.uk/ols/ontologies/atol">https://www.ebi.ac.uk/ols/ontologies/atol</a>
EOL	<a href="https://www.ebi.ac.uk/ols/ontologies/eol">https://www.ebi.ac.uk/ols/ontologies/eol</a>
UBERON	<a href="https://www.ebi.ac.uk/ols/ontologies/uberon">https://www.ebi.ac.uk/ols/ontologies/uberon</a>
CL	<a href="https://www.ebi.ac.uk/ols/ontologies/cl">https://www.ebi.ac.uk/ols/ontologies/cl</a>
BTO	<a href="https://www.ebi.ac.uk/ols/ontologies/bto">https://www.ebi.ac.uk/ols/ontologies/bto</a>
CLO	<a href="https://www.ebi.ac.uk/ols/ontologies/clo">https://www.ebi.ac.uk/ols/ontologies/clo</a>
SO	<a href="https://www.ebi.ac.uk/ols/ontologies/so">https://www.ebi.ac.uk/ols/ontologies/so</a>
GO	<a href="https://www.ebi.ac.uk/ols/ontologies/go">https://www.ebi.ac.uk/ols/ontologies/go</a>
NCIT	<a href="https://www.ebi.ac.uk/ols/ontologies/ncit">https://www.ebi.ac.uk/ols/ontologies/ncit</a>
CHEBI	<a href="https://www.ebi.ac.uk/ols/ontologies/chebi">https://www.ebi.ac.uk/ols/ontologies/chebi</a>

In WP3, BovReg is committed to update these ontologies to ensure that BovReg metadata is accurately described.

#### 2.4. Increase data re-use (through clarifying licences)

As described in section 1 above, there are two classes of BovReg omics data, (1) data made immediately public once generated in accordance with the FAANG data sharing statement and (2) data that will only be available to members of the BovReg consortium through a private data portal until it is later publicly released upon use in a publication or at the end of the project.

The data generated in accordance with the FAANG data sharing principles (<https://www.faang.org/data-share-principle>), is based upon the principles of the Toronto (<https://www.nature.com/articles/461168a>) and Fort Lauderdale (<https://www.genome.gov/Pages/Research/WellcomeReport0303.pdf>) agreements. This reserves the right for BovReg to make the first publication with the data, whether a dataset has an associated publication is tracked clearly in the FAANG data portal (<https://data.faang.org/home>). All FAANG datasets will be clearly labelled with these data sharing principles, with the following statement:

*"This study is part of the FAANG project, promoting rapid prepublication of data to support the research community. These data are released under Fort Lauderdale principles, as confirmed in the Toronto Statement (Toronto International Data Release Workshop. Birney et al. 2009. Pre-publication data sharing. Nature 461:168-170). Any use of this dataset must abide by the FAANG data sharing principles. Data producers reserve the right to make the first publication of a global analysis of this data. If you are unsure if you are allowed to publish on this dataset, please contact the FAANG Consortium and FAANG Data Coordination Centre (email both [faang@iastate.edu](mailto:faang@iastate.edu) and [faang-dcc@ebi.ac.uk](mailto:faang-dcc@ebi.ac.uk)) to enquire. The full guidelines can be found at <http://www.faang.org/data-share-principle>."*

The BovReg private data will initially only be available to consortium partners whilst publications are prepared. Upon publication or by the end of the project the BovReg omics datasets in sections 1.1.A and 1.1.B will be made available through the EMBL-EBI public archives and indexed in the FAANG data portal (<https://data.faang.org/>).

The software developed by the BovReg consortium will be openly licensed for reuse and will display this license in the root folder of the public repository. Data quality assurance processes and metrics will be investigated and implemented as part of the pipeline development. Wherever possible pipeline development will utilise the latest open source, published and recognised analysis software. The software will be available through the FAANG GitHub repository (<https://github.com/FAANG/>).

### 3. Allocation of resources

EMBL-EBI is funded by BovReg to develop, maintain and support the private data hub system for handling of consortium raw and analysis of omics data during the life of the project. It also funds key development objectives to support FAIR data archiving and presentation through the public FAANG data portal, to which the BovReg data will be archived for long term availability. Key objectives of the development include enhancements to the metadata data recording for BovReg samples and the applicability of ontology terms to describe the data. These developments will also be of benefit to the wider FAANG and cattle communities. The institutes that are generating BovReg datasets ensure accurate FAIR recording of sample and experimental datasets that will be validated using the FAANG data portal tools (<https://data.faang.org/validation/samples>).

The FAANG Data Coordination Centre at EMBL-EBI is responsible for coordination of data management for the BovReg project and the development of the BovReg data private platform that manages BovReg data while publications are prepared. The metadata system of the private data hub is purposefully identical to the FAANG public metadata standards used in the FAANG data portal, thus ensuring rapid, FAIR and H2020 compliant data release upon acceptance of relevant publications. Upon publication or by the end of the project data newly generated under the BovReg project will be openly archived in the EMBL-EBI public archives and indexed through the FAANG data portal (<https://data.faang.org/>). The EMBL-EBI archives have separate long term funding for the long-term persistence of the data beyond the end of the project.

## 4. Data security

### Omic Data

BovReg will submit all omics data at the earliest opportunity either to the BovReg private data hub or the public archives (as described in section 1), both of which are hosted at EMBL-EBI.

The BovReg private data hub incorporates the AAP authentication system

(<https://aai.ebi.ac.uk/home>) to ensure access is restricted to BovReg consortium members.

The data system is both secure and incorporates backup across multiple geographically different data centres. The metadata system and underlying infrastructure is based on the same principles as the public FAANG data portal, so that when the BovReg data is needed to be made publicly available it is already in the appropriate format and can be made instantly available in the public FAANG data portal.

As with other FAANG consortium datasets, the BovReg FAANG data will be submitted to EMBL-EBI public archives and catalogued by the FAANG data portal. This process assigns a globally recognised unique identifier for the long-term preservation of the datasets. The EMBL-EBI public archives utilise three discrete Tier III plus data centres in different geographical locations to ensure long-term data security. Research sequencing data is also replicated through the International Nucleotide Sequence Database Collaboration (INSDC; <http://www.insdc.org/>) agreements that ensures the data is replicated at the National Center for Biotechnology Information (NCBI; <https://www.ncbi.nlm.nih.gov/>) and DNA Databank of Japan (DDBJ; <https://www.ddbj.nig.ac.jp/index-e.html>). Each of the three INSDC resources agree to recognise the identifiers assigned by the other members, replicate the data presentations and act as a geographical mirror of all datasets.

Once data is released to the public domain through the EMBL-EBI public archives, EMBL-EBI commit to store the data in the public archives for the foreseeable future, this is part of EMBL-EBI's core commitment to ensure public scientific data remains available through its core data resource archives.

### Personal data

- The interview data collected in Task 8.1 of WP8 will be stored with the help of the Research Data Management Support Office of Utrecht University.

- The Democs game outputs of Task 8.2 of WP8 will be stored at the University of Edinburgh DataShare repository.

## 5. Ethical aspects

### Omics data

The BovReg project has been designed to fully comply with all ethical and legal requirements for the relevant countries in which its partners operate. The omics data is not based on human subjects and thus does not have any legal or ethical issues regarding the sharing of the data it generates, and informed consent is not required for data sharing or storing of personal data.

### Personal data

The only personal data collected will be:

- Interview results from WP8 task T8.1.
- Web logs

BovReg will fully comply with General Data Protection Regulations for its activities and web services, with GDPR statements and terms of use available on each BovReg site page (<https://www.bovreg.eu/>) and data pages on the FAANG data portal (<https://data.faang.org/>). The implementation of GDPR regulations will be detailed in our Protection of Private Data statement, submitted as deliverable D12.3.

## 6. Other issues

The FAANG consortium ([www.faang.org](http://www.faang.org)) has initiated through the Data Coordination Centre at EMBL-EBI via independent funds infrastructural support for the coordination, validation and archiving data generated by FAANG related projects. This infrastructure will be supported throughout the BovReg project in accordance with its data requirements. BovReg was funded under the H2020 SFS-30-2018-2019-2020 - Agri-Aqua Labs programme that dictates in the call that the data generated complies with FAANG metadata standards (<https://data.faang.org/ruleset/samples#standard>) and is archived in the relevant public biological data archives for indexing by the FAANG data portal (<https://data.faang.org/>).