ELIXIR Domestic Animals Genome and Phenome Community RoadMap 2025-2030

Table of Contents

Introduction

The ELIXIR Domestic Animals Genome and Phenome (DAGP) Community will support genotype to phenotype analysis for farmed and companion animal species. As a community we aim to coordinate, discuss and explore the potential for data/technology solutions to address key issues in animal welfare, behaviour, health, infectious diseases, metabolism, nutritional efficiency, and preservation of genetic diversity and the environment. This will consolidate efforts to develop data standards, coordination, workflows and visualisation, which are needed to enhance the science underpinning rapidly developing fields in domestic animal genomics, including pangenome analysis, functional genomics, genome editing, phenotyping, and bio-banking. The developed standards will be based on the FAIR principles (Findable, Accessible, Interoperable, Reproducible) [1], leveraging established data and metadata sharing principles, standards and ontologies to promote best practices in data coordination, analysis and archiving, as well as promoting readiness for use of AI tools.

The RoadMap for the ELIXIR Domestic Animal Genomes and Phenomes Community 2025-2030 builds on the current data infrastructure, resources and tools that are available for domestic animal genomics and phenomics, particularly the global Functional Annotation of Animal Genomes (FAANG) Data Sharing Principles [2] and FAANG Data Portal [3], and provides a set of community-led objectives. These objectives describe how ELIXIR services can be applied in the field of domestic animal genomics to benefit the entire community. Some examples include: i) expanding the FAANG Data Portal to include phenotype data and additional species, in cooperation with the ELIXIR Platforms, ii) exploring submissions of new and linked data types across ELIXIR services, including single cell data, and provision of encrypted proprietary data from industry partners, and iii) strengthening connections to existing ELIXIR Communities, infrastructures and international consortia. We have established strong links, for example, with the Plant Science Community through the similar challenges faced for food and agriculture. Our objectives are designed to contribute directly to the ELIXIR Biodiversity, Food Security and Pathogens and the Cellular and Molecular Research scientific priority areas.

Objectives

Enhance visibility, findability and sustainable access to data for domestic animals: Encourage the use of established archives and databases, utilising current metadata standards such as those established by the FAANG consortium, and develop or adapt standards for animal phenotyping data. Identify valuable resources and explore new opportunities to consolidate and utilise these for improving data management and sharing.

Facilitate data integration and interoperability by providing examples of expected use cases across different data types and platforms e.g. for ingestion of single cell data for domestic animals into the single cell Expression Atlas (scEA).

Deliver containerised bioinformatic workflows for data analysis through strategic links with the nf-core Animal Genomics Special Interest Group.

Ensure long-term sustainability of domestic animal data resources, data portals and databases, including the FAANG Data Portal, by promoting best practices in data and metadata submission, validation, and archiving, and fostering alignment with global, national and European infrastructure initiatives, through securing continuous funding.

Support the development of machine learning and AI-driven approaches, such as large language models (LLMs), for data integration, curation and analysis to improve automation and scalability, and facilitate complex queries of the datasets through existing data portal infrastructure.

Build capacity for sample and data sharing between biobanks by developing a standard based on the frameworks provided by BrAPI in the plant domain or the BBMRI directory for human health contributing to conservation of genetic diversity.

Provide a framework for consolidating data and metadata recording and archiving practices for phenotyping data, through understanding current workflows and preferred deposition databases, and building on existing experience, particularly from the Plant Science Community.

Work with industry stakeholders on standards for measurements that can be applied at the point of data collection to facilitate prioritisation and conversion of the data from farming equipment into measured traits in collaboration with internationally recognised organisations like the International Committee for Animal Recording (ICAR) and the ISO Technical committee 347 for Data-driven agrifood systems.

Build joint activities within and beyond the Domestic Animals Genome and Phenome Community leveraging existing initiatives for domestic animals, such as FAANG, to encourage the adoption of data standards by researchers, data and metadata repositories, and in scientific publications.

Facilitate options for data sharing with industry stakeholders through digital twins, data encryption, anonymisation and/or other methods for sharing proprietary data, such as infrastructure changes to existing resources.

Develop community recommendations, create an open dialogue and contribute to ELIXIR's resources, enabling their broad adoption through targeted training and industry-focused materials to support adoption by data holders and researchers.

Perform gap analysis to identify missing tools and resources to advance the goals of the Domestic Animal Genome and Phenome Community and address them through ELIXIR supported mechanisms.

Strengthen engagement with the Biodiversity, Food Security & Pathogens (BFSP) and Cellular & Molecular Research (CMR) scientific priority areas by actively involving new stakeholders from across communities, and fostering joint activities based on developing shared resources.

Ensure that the Domestic Animal Genome and Phenome Community activities meet the demands and recommendations of researchers through internal communication in the Community and external joint actions, such as through review or opinion publications, and participation in working groups, EU COST-Actions and conferences.

Support the development and sustainability of community-specific tools, data portals, and repositories, promoting their use and ensuring their continued relevance, usability, and integration within the ELIXIR concept.

Implement this roadmap through European and global collaborations, actively contributing to international projects, through coordination and delivery of activities, and securing their funding.

RoadMap Work Plan Overview

Over the coming five years we have six main tasks that encompass the objectives listed above. These are described below with some examples of how they could be achieved.

Task 1 Data Standards, Development and Dissemination

The increasing volume, complexity, and diversity of data sets in fields such as animal genomics and phenomics, along with the rise in open data and open science initiatives, necessitate standardised and comprehensive data descriptions. These descriptions are essential for the findability, retrievability, and interoperability of data and must be documented during the data generation process. The Domestic Animal Genome and Phenome Community has endorsed existing descriptions from previous and ongoing projects, such as the FAANG metadata standards [2] for omics datasets, and the Vertebrate Breed Ontology [4]. Additionally, we aim to enhance the descriptions for phenotyping metadata. To achieve this, we are collaborating closely with the Plant Science Community to align with their commonly used ISA (Investigation-Study-Assay [5], standardisation efforts in the upcoming BioHackathon [biohackathon-projects-2025/17.md at main · elixir-europe/biohackathon-projects-2025 ·

GitHub], as well as working with the biodiversity community to adapt the ABCD standard for farm animals [Use Cases – Our Pilot Projects - NFDI4Biodiversity]. To provide more semantic meaning mapping to other important linked data resources, such as AGROVOC [6] and continuously growing ontologies like ATOL, EOL, and AHOL (https://bioportal.bioontology.org/ontologies/ATOL?p=classes&conceptid=root) are also incorporated. Also, when considering automatic data exchange, the Global standard for livestock data ICAR, and especially the ADE (Animal data exchange) (https://github.com/adewg/ICAR), which is currently under development, should be considered for coming standardisation efforts. Eventually, this will be applicable to the databases for gene banks to go beyond data sharing towards sharing of biological samples contributing to conservation of local and national genetic diversity.

Task 2 Data Collection

To facilitate easier ingestion of data into the archives we are working with data producers to develop standards that can be applied at the point of data collection. For example, we are working with industry stakeholders using ISO/TC 347 Datadriven agrifood systems to take measurements from farming equipment (fat content, amount of milking each time a cow is milked etc) and convert them into measured traits like milk yield based on standards developed by breeding organisations and ICAR. On farm data often includes 100's of 1000's of data points and images collected over time from sensors or cameras, e.g. for observational time series data for behaviour studies and feed intake measured using automated feeders. Image data is often very large, and not all the data is sufficiently informative to require archiving. As such well-defined data standards established at the point of data collection make downstream prioritisation of the data much easier.

Task 3 Data Discoverability

The FAANG Data Portal [3] currently supports the entire domestic animal genomics community, through open FAIR sharing of data, complete with standardised rich metadata. FAANG projects produce a standardised set of multi-omic assays with resulting data placed into ELIXIR open data archives. To ensure this data is easily findable and accessible by the community, the portal automatically identifies and collates all submitted FAANG data into a single easily searchable resource. The portal also has the flexibility to incorporate new technical infrastructure to effectively deliver new data types and technologies to best fit the needs of the domestic animal science community. As a community, we plan to further explore options to expand the capacity of the FAANG Data Portal to deliver data for emerging species, including insects, encrypted data and phenotype data. In the longer term, large language models and other machine learning tools will allow much more complex dynamic data queries which will greatly enhance the discoverability potential provided by the portal infrastructure.

Our aim over the coming five years is to ensure the sustainability of the FAANG Data Portal through funding applications to the European Commission, and through shared goals and objectives with the global FAANG initiative [7].

Task 4 Data Integration and Interoperability

One example of how we plan to maximise interpretability and reuse of domestic animal datasets, is to integrate them with model-organism resources, using zebrafish as a case study. Specifically, we will (i) harmonise metadata and ontologies with FAANG and incorporate zebrafish genomic and epigenomic resources, namely DANIO-CODE (https://danio-code-dcc.genereg.net/, [8]) data into the FAANG data portal; (ii) leverage orthology and synteny frameworks (e.g., Ensembl Compara [9] and Genomicus [10]) to transfer gene- and enhancer-level annotations across species; (iii) perform comparative non-coding genome analysis, including inference of conserved regulatory programs with information-preserving projection (IPP; [11]); and (iv) integrate multi-species single-cell atlases (embryogenesis, tissue regeneration) developed within DANIO-ReCODE (https://danio-recode.eu) to map cell-type and state correspondences. This integration will enable zebrafish to serve both as a comparative anchor and as an experimental validation system where transgenesis in domestic species is impractical (e.g., long generation times, seasonal breeding). The approach aligns with ELIXIR best practices for FAIR metadata, interoperable vocabularies, and reproducible workflows—ensuring that cross-domain resources can be queried jointly for hypothesis generation and validation. Another example we plan to pilot to maximise data integration is a homomorphic-encryption workflow using sensitive, proprietary, salmon aquaculture phenotype/trait data to test practical encrypted submission and joint G2P analysis via the FAANG Data Portal and evaluate the improvements in genetic testing with larger integrated datasets, all while protecting proprietary records.

Task 5 Data Analysis

To guarantee that analysis results remain a long-term asset to the scientific community, it is essential that they are produced using bioinformatics pipelines that adhere to the FAIR principles. The nf-core community have established a set of standards to deliver reproducible and interoperable data analyses (FAIR) pipelines implemented with Nextflow as a workflow management system [12]. Implementing these standards ensures that computational workflows are not only reproducible and transparent but also facilitate seamless integration with other research efforts so that comparative analysis can be performed. In the framework of animal research, the EuroFAANG consortium, which works on decoding genotype-to-phenotype relationships of farm animals, has shown that the adoption of nf-core as its standard for analysis workflows has been instrumental in leading to more efficient and reproducible cross-institutional collaboration and analysis [13]. Based on this we plan to continue our collaboration with the nf-core community by leveraging their extensive library of modules, subworkflows and pipelines to progressively adopt FAIR principles for workflows. Within nf-core, there is an Animal Genomics Special Interest Group (https://nf-co.re/special-interest-groups/animal-genomics) and a dedicated Slack channel, which serve as key platforms for collaboration. Additionally, the Domestic Animals Genome and Phenome Focus Group has already established spaces to foster interaction between nf-core members and the animal genomics and phenomics community. The goal is to ensure that nf-core pipelines can be continuously improved to meet the specific needs of animal genomics / multi-omics and phenomics analyses, facilitating the development of high-quality, standardized workflows for animal research.

Task 6 Collaboration

Members of the community are involved in many European and global initiatives for domestic animals including the GenoPHEnix ESFRI proposal (https://genophenix-ri.eu) to build consolidated infrastructure for farmed animal science and biobanking in Europe. We also participate in international committees dedicated to developing standards for animal recording such as ICAR and have strong links with consortia aiming to capture local and global genomic diversity such as the Ruminant T2T project [14] and Bovine PanGenome Project [15]. With representation on the steering committees of large-scale genomics initiatives for domestic animals such as the FAANG [16] and FarmGTEx [17] projects we are ideally placed to ensure the needs of the Domestic Animals Genome and Phenome Community are represented globally. Over the coming five years we hope to facilitate training workshops and develop shared objectives with the Animal Breeding - RCN: Farm Animal Genomics Collective in the US (https://portal.nifa.usda.gov/web/crisprojectpages/1032104-animal-breeding-rcn-farm-animal-genomics-collective.html) and with the EU-LI-PHE COST-Action focused on livestock phenomics (https://eu-li-phe.eu/). To deepen engagement with industry, we will organize ELIXIR Industry Engagement Day events, starting with the Aquaculture Industry Engagement Day (October 2025), that convenes salmon producers and breeding companies with ELIXIR/EuroFAANG partners to set joint priorities.

Acknowledgements

The authors would like to thank all members of the Domestic Animals Genome and Phenome Focus group, the ELIXIR hub, and particularly the Plant Science and Biodiversity Communities for input into the development of this RoadMap.

References

1. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018. https://doi.org/10.1038/sdata.2016.18

2. Harrison PW, Fan J, Richardson D, Clarke L, Zerbino D, Cochrane G, et al. FAANG, establishing metadata standards, validation and best practices for the farmed and companion animal community. Anim Genet. 2018;49:520–6. https://doi.org/10.1111/age.12736

3. Harrison PW, Sokolov A, Nayak A, Fan J, Zerbino D, Cochrane G, et al. The FAANG Data Portal: Global, Open-Access, "FAIR", and Richly Validated Genotype to Phenotype Data for High-Quality Functional Annotation of Animal Genomes [Internet]. 2021. https://www.frontiersin.org/article/10.3389/fgene.2021.639238

4. Mullen KR, Tammen I, Matentzoglu NA, Mather M, Balhoff JP, Esdaile E, et al. The Vertebrate Breed Ontology: Towards effective breed data standardization. United States; 2025.

5. Rocca-Serra P, Brandizi M, Maguire E, Sklyar N, Taylor C, Begley K, et al. ISA software suite: supporting standards-compliant experimental annotation and enabling curation at the community level. Bioinformatics. 2010;26:2354–6. https://doi.org/10.1093/bioinformatics/btq415

6. Subirats-Coll I, Kolshus K, Turbati A, Stellato A, Mietzsch E, Martini D, et al. AGROVOC: The linked data concept hub for food and agriculture. Comput Electron Agric. 2022;196:105965. https://doi.org/10.1016/j.compag.2020.105965

7. Clark EL, Archibald AL, Daetwyler HD, Groenen MAM, Harrison PW, Houston RD, et al. From FAANG to fork: application of highly annotated genomes to improve farmed animal production. Genome Biol. 2020;21:285. https://doi.org/10.1186/s13059-020-02197-8

8. Baranasic D, Hörtenhuber M, Balwierz PJ, Zehnder T, Mukarram AK, Nepal C, et al. Multiomic atlas with functional stratification and developmental dynamics of zebrafish cis-regulatory elements. Nat Genet. 2022;54:1037–50. https://doi.org/10.1038/s41588-022-01089-w

9. Dyer SC, Austine-Orimoloye O, Azov AG, Barba M, Barnes I, Barrera-Enriquez VP, et al. Ensembl 2025. Nucleic Acids Res. England; 2025;53:D948–57. https://doi.org/10.1093/nar/gkae1071

10. Nguyen NTT, Vincens P, Dufayard JF, Roest Crollius H, Louis A. Genomicus in 2022: comparative tools for thousands of genomes and reconstructed ancestors. Nucleic Acids Res. 2022;50:D1025–31. https://doi.org/10.1093/nar/gkab1091

11. Phan MHQ, Zehnder T, Puntieri F, Magg A, Majchrzycka B, Antonović M, et al. Conservation of regulatory elements with highly diverged sequences across large evolutionary distances. Nat Genet. 2025;57:1524–34. https://doi.org/10.1038/s41588-025-02202-5

12. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020;38:276–8. https://doi.org/10.1038/s41587-020-0439-x

13. Langer BE, Amaral A, Baudement M-O, Bonath F, Charles M, Chitneedi PK, et al. Empowering bioinformatics communities with Nextflow and nf-core. Genome Biol. 2025;26:228. https://doi.org/10.1186/s13059-025-03673-9

14. Kalbfleisch TS, McKay SD, Murdoch BM, Adelson DL, Almansa-Villa D, Becker G, et al. The Ruminant Telomere-to-Telomere (RT2T) Consortium. Nat Genet. United States; 2024;56:1566–73. https://doi.org/10.1038/s41588-024-01835-2

15. Smith TPL, Bickhart DM, Boichard D, Chamberlain AJ, Djikeng A, Jiang Y, et al. The Bovine Pangenome Consortium: democratizing production and accessibility of genome assemblies for global cattle breeds and other bovine species. Genome Biol. 2023;24:139. https://doi.org/10.1186/s13059-023-02975-0

16. Giuffra E, Tuggle CK and. Functional Annotation of Animal Genomes (FAANG): Current Achievements and Roadmap. Annu Rev Anim Biosci. 2019;7:65–88. https://doi.org/10.1146/annurev-animal-020518-114913

17. Fang L, Teng J, Lin Q, Bai Z, Liu S, Guan D, et al. The Farm Animal Genotype–Tissue Expression (FarmGTEx) Project. Nat Genet. 2025;57:786–96. https://doi.org/10.1038/s41588-025-02121-5