CoBreeding Source Code DMP

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - DMP template (english)" fourni par Agence nationale de la recherche (ANR).

Plan Details

Plan title	CoBreeding Source Code DMP
Deliverable	D0.1
Version	First version
Fields of science and technology (from OECD classification)	Agriculture, forestry, and fisheries
Language	eng
Creation date	2023-03-29
Last modification date	2023-06-05

Project Details

Project title	Co-design of animal and plant breeding schemes with multi-performance objectives (economic, social and environmental) to develop agroecological production
Acronym	CoBreeding
Abstract	New ways of mobilizing genetic diversity must be explored to respond better and faster to the challenges of the agroecological transition, climate change (adaptation, mitigation), global health and animal welfare, environmentally friendly plant cultivation practices, and consumer expectations. Breeding programs need to evolve so that larger numbers of diverse and complementary genotypes become available to fit the full range of current and future environmental conditions and production systems. To develop and implement new genetic improvement strategies, systemic approaches are needed to revisit the technical and socio-economic elements that determine the choices of populations/breeds/lines/varieties and individuals and their orientations in the different animal and/or plant production systems. The main levers are (i) the development of “co-design” approaches including the various actors concerned by these production systems to favor disruptive innovation for agroecology for building new social agro-ecosystems and defining innovating living entities ,(ii) the ability to evaluate multi-performance objectives (economic, social, environmental), (iii) diversification to improve the overall resilience of production systems, that imply to design new breeding strategies, but also new populations and farming practices, and to better characterize

genotype x environment x farming practices interactions. All these issues are common to the genetic improvement of animals and that of cultivated plants and will advantageously be addressed together in the CoBreeding project. To meet these challenges, the CoBreeding project is structured around three main axes. Axis A will put the emphasis on the process of collective design, considering heterogeneous actor communities. Axis B will concern the enrichment in multicriteria genetic evaluation methodologies, including environmental complexity. Axis C will focus on diversification and breeding strategies. By combining the results of experiments and field data, newly developed digital methods based on mathematics and computer modeling will highlight the synergies and antagonisms between the technical, economic, environmental and societal components of a system's performance. They will be used to optimize production systems and implement case studies for developing innovative lines/cultivars and associations for selected animal and plant species. Some data collection funded in project I.2 will be required to get the proofs of concept of the feasibility and impact of our innovations in breeding scheme design. We will also benefit from resources collected in flagship projects II.1, II.2 and infrastructures Liph4SAS and AGROECOPHENO to collect new genotypes and phenotypes. The objectives will be to co-produce knowledge and generic tools facilitating efficient and dynamic genetic management of livestock populations and crops, with the participation of selection operators, agricultural technical services, and direct or indirect users of the innovations produced (farmers, consumers, citizens) to meet the objectives of agroecological transformation. The CoBreeding project requires a funding of 3 M€ and brings together over 80 recognized specialists in biology (genetics, animal and plant sciences), data sciences (bio-informatics, mathematics, statistics) and social sciences from INRAE, INRIA, AgroParisTech, AgroCampus-Ouest, Université Paris-Saclay, Mines Paris - PSL, ENSFEA and Oniris who will devote 360 working months of permanent staff to the project.

Funding

- Agence Nationale de la Recherche : ANR-22-PEAE-0003

Start date 2023-01-01

End date 2027-12-31

Partners

- université Paris sciences et lettres https://ror.org/013cjyk83
- Institut national de recherche en informatique et en automatique https://ror.org/02kvxyf05
- Institut national de recherche pour l'agriculture, l'alimentation et l'environnement https://ror.org/003vg9w96
- Institut des sciences et industries du vivant et de l'environnement (AgroParisTech) https://ror.org/02kbmgc12
- Université Paris-Saclay https://ror.org/03xjwb503

Research outputs :

1. Source code output (Software)

Contributors

Name	Affiliation	Roles
De Oliveira Yannick		• DMP manager • Personne contact pour les données
Phocas Florence - 0000-003-1161-3665	INRAE	• Project coordinator

CoBreeding Source Code DMP

1. Data description and collection or re-use of existing data

1a. How will new data be collected or produced and/or how will existing data be re-used?

Source code can be produced using an IDE (Intergrated Development Environment) such as Eclipse or RStudio for R scripts or another one. Simple source code editor can also be used.

Existing code can be reused if licences are compatible with open source licences.

Task A2b will be developed froom two existing softwares developed at INRAE, SiriusQuality and PigOptim.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Source code are mainly text files with extension regarding the langage that is used (.py for python, .java for Java .R for R scripts etc.)

Volume is less than 1Mo often.

---

2. Documentation and data quality

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

No specific metadata standard is used for Source Code.

But a documentation will be provided for each software. It can be a README file explaining how to use the script.

For complexe software, a full documentation will be provided. For instance, documentation of DiverCILand will be hosted on readthedocs.org plateform.

2b. What data quality control measures will be used?

Source Code will be hosted on the INRAE GitLab instance (https://forgemia.inra.fr). Git versionning control system will be used to keep versions of scripts and code source files.

---

3. Storage and backup during the research process

3a. How will data and metadata be stored and backed up during the research?

Source code are hosted on local computers for development and stored on the INRAE GitLab instance.

This instance is backed up by INRAE MIA IT Team.

3b. How will data security and protection of sensitive data be taken care during the research

The GitLab instance of INRAE can be accessed with LDAP credential. Source code can be freely cloned but contributing to the source code can be controlled by administrator of the repository.

4. Legal and ethical requirements, code of conduct

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

No personnal data

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

All source code will be available using open source licences (GPL, AGPL etc.) reagrding the nature of the source code.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

None

5. Data sharing and long-term preservation

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Source code will be shared as soon as possible during or after the project. Source code will be available on the GitLab instance of INRAE. In some cases a mirror can be used on the most popular forges such as GitHub.com ou GitLab.com
Source code can be published with other types of data if it is relevant.

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data repository or archive)?

Source code will be hosted on INRAE GitLab instance as long as this service exists.

5c. What methods or software tools are needed to access and use data?

Source code do not need specific software to be read. Anyway to be executed source code need specific interpreter (R, python etc.).

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

Question sans réponse.

---

6. Data management responsibilities and resources

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

People in charge of the code source management are the administrator of the project in INRAE GitLab instance.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Question sans réponse.