Exchanges of French genetic data are driven by farmer consent managed in a dedicated blockchain

L. Journaux[1], L. Duveaux[2], H. Gilles[3] and S. Picardat[4]

[1]*France Génétique Elevage (FGE), 149 rue de Bercy 75595 Paris cedex 12, France*
Corresponding Author: laurent.journaux@france-genetique-elevage.fr
[2]*Association SPIE, 9 avenue George V, 75008 Paris, France*
[3]*FAST, 149 rue de Bercy 75595 Paris cedex 12, France*
[4]*Agdatahub, 9 avenue George V, 75008 Paris, France*

Abstract

FGE manages French collective system for the genetic improvement of ruminants and offers services to its members with shared genetic information systems. The information system for cattle genetic improvement manages 200 million of animal (+7.5 million per year) and more than 1 billion of information (+40 million per year), coming from 44 000 farms, recorded and used by more than 300 companies (135 with a very significant activity).

FGE launched in 2020 the deployment of a new professional national data base, connected with a dedicated data exchange platform and a breeder's consents verification platform "Agata Consent" to validate and secure the exchange based on a consortium blockchain solution. The system, in production for the first users, ensures the informed consent of breeders speaking about the use of their data by several professional organizations for research and development, genetic improvement or innovation, without any intervention from the user. It works as a trusted third party to exchange data even with multiple addressing, with addressing parameters supported by web services. It increases the confident of breeders and organizations in the exchanges and the uses of data all along the value chain, in adequation with the European Data Act regulation.

Introduction

Since the 60's, the French cattle breeders' organizations have been structuring their technical data recorded on farm for ruminant production (cattle, sheep and goat) at regional and national levels. Initially, the investment and infrastructures were funded and performed by state.

Gradually, the part of investment and cost supported by breeder's organization increased. Today, the cost and investments of the system are supported by farmer's funds for 90% and by state subsidies for 10%.

A first step occurred in 1998 to renew all the system in a distributed relational database. The system was well adapted to the land organization with, small to medium size, local farmer associations or cooperatives (Rehben *et al.*, 2005; Rognant *et al.*, 2005). The system, currently in production (Balvay 2017, Balvay *et al.,* 2014), has been managed by the national professional organization France Génétique Elevage (FGE) for 14 years with the development of new tools as consent management.

For Cattle, the information system covers animal identification, parentages, performances (i.e. milk production, weight, fertility and so on) and genotyping (DNA information) with all the technical information to use this data for statistical treatments (location, birth date, calving rank…). These data are specially used to estimate the breeding value of animal and to perform advices for the farmers; these data are exchanged at international level to establish international ranking of artificial insemination sires on breeding values. In addition, this information is available for farmers themselves and others clients like private companies.

The amount of data collected each year is very large as the total amount of information cumulated for 60 years.

The number of actors who contribute to the system is large: more than 300 data providers in 2019. 135 of them exchanged more than 10 000 records.

One specificity of this information system is to manage data on a very long term: the statistical studies performed to estimate breeding values needs a very large number of data and a knowledge of pedigree with a maximum of depth (minimum 4 generations of ancestors, i.e. data managed during a minimum of 30 years for Cattle!)

Table 1. Number of data managed in the cattle information system (in Millions)

	Identities	Births	Lactations	Milk records	Weighing	Morphology records	Inseminations
Records created in 2021	7,012	6,749	2,232	20,393	1,271	1,156	6,514
Total in the database	233,6	158,0	114,3	857,0	49,1	25,2	189,0

A new landscape for regulation

The data exchange in France are regulated by EU regulation on personal information – RGPD, CNIL 2022. In the next year this regulation will be updated by the Data Governance Act. While the Data Governance Act, presented in November 2020 and agreed by co-legislators in November 2021, creates the processes and structures to facilitate data sharing by companies, individuals and the public sector, it also clarifies who can create value from data and under which conditions. Especially, the Data Governance Act defines a new role of Data Intermediary Service Provider which must be independent of any storage or treatment of any kind. Furthermore, to ensure the liability and accessibility of data from production, beneficiary companies will have to ask for an explicit consent (prolonged GDPR for non personal data). This consent must be also verifiable on behalf of any data exchange. This new regulation will be in force by the end of 2023 (EU, 2022).

A new landscape of actor for new needs on data

Nowadays, the situation has evolved a lot compared to the beginning of the century. In some parts of the territory some small organizations remain structured around regional information systems managed by breeder cooperatives or associations dedicated to this technical activity. But, in other parts, big players (more than 100 M€ income) are individually part of the game. They have an internal information system covering all

the aspects of the management of the company, technical data dedicated to genetic improvement is now only a part of information system structured in ERP (Enterprise Resource Planning).

The demand for exchange of data for multiple use is now very frequent and push the development of new solutions of exchange or sharing of data (iDDEn Reents and Pekeler, 2021*;* DataGene, Shaffer 2021*;* US Dairy Brain project Cabrera *et al.*, 2021.

Since 2017, FGE engaged a large project to adapt the existing system to this new deal. After 3 years of initial studies the project is entering into the final phase of development and production.

Description of the new system

The new system is a national database managed by FGE to capitalize all the data collected since the 60's and a set of web services to support data exchange with all the regional or private information systems.

The architecture of the new information system is organized to combine a national central data base with regional platform shared by several actors and private information systems where the technical information managed for genetic purposed is a part of integrated ERP of the company.

To take into account the mandatory respect of breeder consent in accordance with French and Eu regulation and to be compatible with policy of French Farmers Union, this new system is combined with a consent management system open to all the companies involved in cattle genetic improvement. This system is based on blockchain: this technology has been chosen for multiple reasons included its capacity of disintermediation and trust creation and the legal value of data encrypted and stored in the blockchain (as described in Leporcher, Goujon, Chouli, 2019)

Exchange between the different databases will be performed by web services

The development of a consent management system paired with a web services exchange system aims at ensuring the informed consent of breeders speaking about the use of their data by several professional organizations for research and development, genetic improvement or innovation. This system is implemented in the new organization to increase the confident of breeders and organization in the exchange and the uses of data.

- If an organization of genetic improvement environment obtained a consent of the farmer for a family of data, it can use the data in accordance with the use stated in the consent and it will have an easy access to the data using the tools developed.

- If a company outside the genetic improvement frame obtained a consent of the farmer and organization who recorded the data's, it can use the data in accordance with the use stated in the consent and it will have an easy access to the data using the tools developed.

The web service solution used is a very standard solution. The publication of web service messages will allow a large use by actors of cattle genetic improvement and other players.

To facilitate data exchange, and extension at national or European level, FGE seeks to maintain the highest compatibility with international data exchange standard. It's Includes, at the first place, the Animal Data Exchange Standard promoted by ICAR (ICAR ADE, 2022). This new French genetic data system already applies all the legal requirements concerning the Data Government Act.

Focus on Base Pro and exchange system

The Base Pro database is based on the PostgreSQL opensource technology.

Base Pro will allow data exchanges between all actors involved in ruminant breeding sector and extended to the livestock value-chain according to modalities compliant with IT state of the art: Web Services.

These tools with standardized interfaces increase the efficiency of access to data by economic players and ensure compliance with management rules for exchanges. In particular, integration into web services of the systematic consent verification before any access to data to check the existence of the data's right holder's consent is a key factor in the confidence and appropriation of the system by these actors..
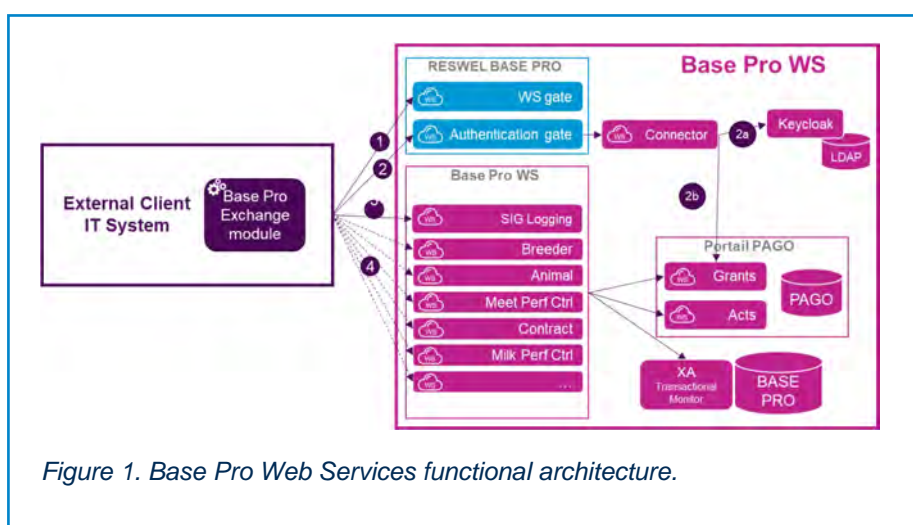

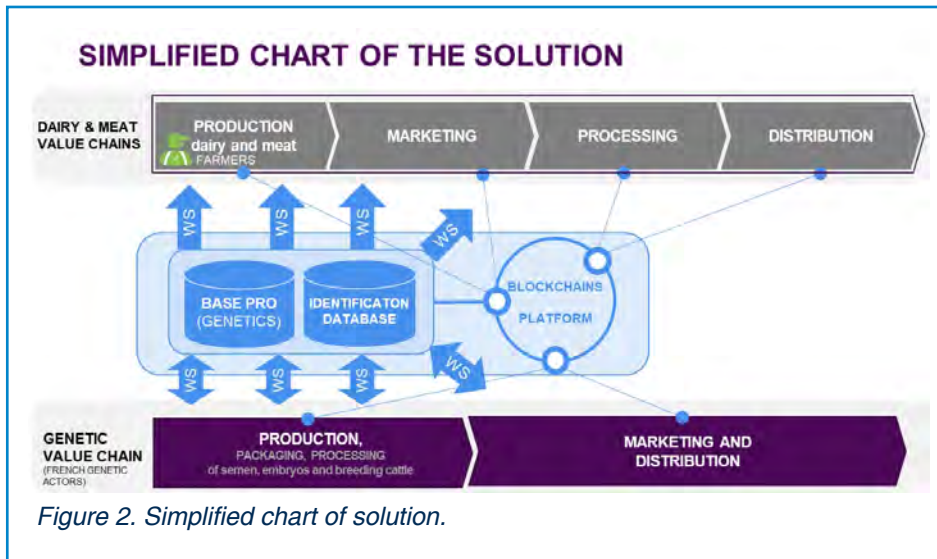*Figure 1. Base Pro Web Services functional architecture.*

In the target architecture the client will access to the web services through the data exchange system (SED) as a trusted third party. The authentication is based on a Keycloak standard solution.

The technical application accounts are stored in a LDAP and managed through the administration portal PAGO which allows the management of the accounts and their access grants.

The technical framework is based on the standard technologies java/Rest/Wildfly/ Swagger.

The Base Pro web services are deployed on a multi-node infrastructure which offers following features:

- Scalablity / Redundancy.
- Security.
- Multi-node load balancing and resilience.
- Continuous integration platform.
- Industrialized platform based on a standard components catalogue.
- Multi-version.

Figure 2. Simplified chart of solution.

- Documented Interface Contract (Swagger).
- Volumetric capacity:
  - 10 000 000 requests/month.
  - Peak : 20 000 000 requests/day.
  - Capacity Load Tests: 13 000 000 requests/hour.

Focus on consent management system

The consent management system called Agata Consent, developed by FAST company (FAST, 2022), was designed to achieve objectives of the Base Pro project: being able to ensure the collection of farmers' consent for data exchanges between various actors of the breeding value chain and giving them access to the consents they gave.

The system is combined with a data exchange management system (SED) developed by Agdatahub company (Agdatahub, 2022) to ensure security and trust.

Administrators of data domains / perimeters can access an ergonomic interface to administrate, create, modify the consents. They can also directly use the available APIs for the different functionalities. Thanks to a detailed rights gestion, farmers access the same interface to consult and accept / refuse consents.

Farmers have their own simple and pragmatic interface to consult and modify (according to contractual parameters) the delegations they gave to professional organizations and the related consents. This interface is exposed on Agri Maker (Agri maker, 2022), a digital services platform for agriculture, already gathering about 33 000 farmers and 25 services.

Agata Consent is a decentralized consent management solution based on open-source components that can be hosted on different cloud platforms using Kubernetes orchestrator. Webservices are developed based on Opensource Springboot framework. Full web management interfaces are developed with Javascript Angular framework,

they are design responsive. Types encoding uses UTF-8 for exchange and storage, dates are encoded using the ISO 8601 norm.

Agata has a distributed architecture with a variable number of nodes, which is coherent with blockchain philosophy and ensure the security of the system by pseudonymization. Agata ensures consents privacy by encrypting all personal data in the blockchain.

Each node of Agata Consent Blockchain has the following technical architecture:

- HyperLedger Fabric Ledger*: data storage in the blockchain.
- Smart contract*: Java code used in the blockchain.
- Couch DB*: database technology Fabric-CA*: certification authority.
- Orderer*: Hyperledger control.

The solution has a Tier III availability classification (with a maximum of 1,6 hours of unavailability per year)

If a specific organization obtained a consent of the farmer for a category of data (also called "data family"), it can use the data in accordance with the usage established in the consent. Moreover, the organization will have an easy access to the data using the tools developed. The system can also be used by a company outside the genetic improvement sector. Indeed, the Consent Management System is generically designed to be able to manage consents relating to data from all sectors of agri-food value chains and even other sectors of economy. It is organized into functional areas called "domains" within which data concerned and rules for managing consents are defined. This organization, which is both modular and generic, makes it possible to offer the service to a wide range of customers, while preserving specificities of each.

Agata Consent has been identified as a mean to obtain the label "Data Agri" for the companies who will use this solution.

For the system initialization, 273 071 existing consents have been loaded. It's concern 74 534 herds and 138 companies.

Agata Consent is technically compatible with the French consent portal AgriConsent, (Agriconsent, 2022) the commercial version of Multipass project (Lauga *et al.*, 2019)

Conclusion

The first 2 domains on which the system combining SED and SGC will be deployed are ruminant identification and genetic improvement. Moreover, even if the system has been developed in this objective, it is applicable to all the agricultural value chain.

The initialization of the system ended in May 2022 and we are now int the phase of increase in load.

The promoters of this solution hope that it will be applied on a large fields of animal data exchange.

Reference

Agdatahub, 2022, https://agdatahub.eu/

Agriconsent, 2022,, https://agri-consent.eu/

Agrimaker, 2022, https://www.agri-maker.com/home.

Balvay B., 2017, A computerized consent management tool forbreeders: why, how? 1997. Proceedings of the ICAR Conference held in Edinburgh (UK) on 14-16 June 2017. M. Winters, M. Coffey, H. Newby, M. Burke and B. Wickham (Eds). ICAR Technical Series no. 22 (ISBN: 92-95014-18-9), p7-12

Balvay B., Rognant R. and Saunier D., 2014, new trends in data capture and exchange solutions between farms and external partners, French situation, ICAR Berlin; Friday May 23th 2014

Cabrera V.E., Fadul-Pacheco L., Wangen S., da Silva T., Zhang F., Fourdraine R.H. and Mattison J., 2021, Collecting, integrating, harmonizing and connecting data from dairy farms: the US Dairy Brain project experience, ICAR Technical Series #25, Circular farming and its impact on animal genetics, animal recording of data and cattle / herd management. Proceedings of the 44th ICAR Conference, virtually held from Leeuwarden, NL, 26-30 April 2021, Editors: G. de Jong, K. de Koning, H. van den Bijgaart. M. Burke and C. Mosconi, September 2021, p 215-219

CNIL, 2022, https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-onjn b

Data Agri, 2022, https://www.data-agri.fr/

EU 2022, European Commission, Data Act: Commission proposes measures for a fair and innovative data economy, 2022, https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113; https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767

FAST, 2022, https://www.fast-agritech.com/

ICAR ADE, 2022, https://www.icar.org/index.php/technical-bodies/working-groups/animal-data-exchange-wg/the-icar-ade-standards/

Lauga B., Balvay B., Topart L., Leclaire J., Clenet A., Brun F., Pinet F., Roussey C. and Sin M., 2019. MULTIPASS: managing the consents of access to farm data in a chain of trustto make new services emerge for farmers, Proceedings of the ICAR Conference held in Prague, CZ, 17-21 June 2019, New traits and adding new value to the recording and id services in the animal production Editors: J. Kucera, P. Bucek, D. Lipovsky, X. Bourrigan and M. Burk ePublisher: ICAR, 2019, p 177-184

Leporcher, Goujon, Chouli, 2019, Les Blockchains de la théorie à la pratique, de l'idée à l'implémentation, collection Epsilon, p18 and 115-120

Rehben E., Balvay B., Béréziat A., Bertrand C., Boulesteix I., Champy R. Dos C., Fusai B., Le Gal H, Marguin L., Poisnel E., 2005, Bilan de la rénovation du système d'information génétique français, Renc. Rech. Ruminants, 2005, 12, p 339-342

Reents R. and Pekeler T., 2021, iDDEN: An international perspective, ICAR Technical Series #25, Circular farming and its impact on animal genetics, animal recording of data and cattle / herd management. Proceedings of the 44th ICAR Conference, virtually held from Leeuwarden, NL, 26-30 April 2021, Editors: G. de Jong, K. de Koning, H. van den Bijgaart. M. Burke and C. Mosconi, September 2021, p 209-212

Rognant R., Breuil G., Gain G., Rehben E., Rouby A., Trocmé D., 2005, Les bases de données collectives d'élevage multi-partenaires: leurs apports aux systèmes d'information des éleveurs et des filières,, Renc. Rech. Ruminants, 2005, 12, p 319-322

Shaffer M., 2021, DataGene, Annual Update 2020/21, https://datagene.com. au/sites/default/files/inline-files/Datagene%20Annual%20Update%202021%20-%20 FINAL.pdf, 28 pages