

e-LivestockProv: An Architecture based on Provenance to Manage Traceability in Precision Livestock Farming¹

Jonas S. Gomes¹, José Maria N. David¹, Regina Braga¹, Wagner Arbex², Bryan Barbosa¹, Wneiton Luiz Gomes², Leonardo M. Gravina Fonseca²

¹Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora, MG – Brazil

²Empresa Brasileira de Pesquisa Agropecuária – Embrapa, Juiz de Fora, MG – Brazil

jonas.gomes@estudante.ufjf.br, {jose.david, regina.braga}@ufjf.edu.br, bryan.barbosa@ice.ufjf.br, {wagner.arbex, wneiton.gomes, leonardo.gravina}@embrapa.br

Abstract. *The use of sensors in the agricultural sector generates a large volume of heterogeneous data that must be processed, stored, and analyzed to support decisions. In addition, decisions taken in agriculture need to be traceable due to the diversity of data and devices present in different agricultural contexts. With provenance, we can trace and analyze data to improve future decisions and avoid the usefulness ones. This article presents the e-LivestockProv architecture, focusing on data provenance.*

1. Introduction

Data accuracy is one of the main goals of using IoT in Digital Precision Livestock. So, farmers are increasingly using sensors to enhance accuracy and monitor food production, animal health, and welfare (Zhai et al., 2020). These sensors generate a large volume of heterogeneous data that needs to be processed, stored, and analyzed by intelligent applications. Based on this data, farmers can extract information and make strategic decisions. Combining data collection and processing can provide accurate decision-making (Karthick et al., 2020) and allow auditable decisions. However, it brings some challenges such as reproducibility, authenticity, and tracking data generated by sensors in smart agriculture. Therefore, it is necessary to analyze and build quality models to guarantee the reliability and quality of the data.

The provenance of data helps in the analysis and tracking of production and animal welfare data and data related to the environment, sustainability, and economics (Bahlo et al., 2019). The use of techniques related to data provenance can help in data traceability, allowing to manage the context of data production, use, and analysis (Buneman, Khanna and Tan, 2001). We believe that by exploring the relationships between data, considering its provenance, combined with intelligent analysis techniques, we will have more efficient support in decision-making in Agriculture 4.0.

Initially, an architecture for collecting, processing, and integrating data in Precision Livestock (see Section 2), named e-Livestock, was proposed (Gomes et al.,

¹ Open peer review artifacts for this paper are available at <https://zenodo.org/communities/opensciense2021>

2021). However, this version did not use provenance for data traceability. Thus, in this work, a new version of the e-Livestock architecture, called e-LivestockProv, is presented, focusing on data traceability and its analysis. Some works have been identified in the literature (Da Cruz et al., 2018) and (Da Cruz et al., 2019), focusing on data provenance. However, they do not deal with issues related to the availability, reliability, authenticity, and diversity of agricultural data that need to be integrated and derived considering their origin, which is our aim.

This paper is organized according to the following structure. Section 2 presents the background and related work. Section 3 presents the proposed architecture, considering the focus on data provenance. An evaluation of the proposal is also presented in this section. In Section 4, final considerations and future work are discussed.

2. Background and Related Work

Internet of Things (IoT) is related to the connectivity of "objects" through the Internet without human intervention. Its use in agriculture and livestock has several advantages, given the possibility of monitoring and controlling agricultural production in a scalable and automated way (Villa-Henriksen et al. 2020). Precision Livestock, or Precision Livestock Farming (PLF), aims to efficiently increase production, improve animal and human welfare, use technologies, targeted resources, and precise control of the production process (Banhazi et al. 2012). In this context, the infrastructure, called Compost Barn, was developed to support the PLF. It aims to reduce maintenance costs of milk production, improve the production and health of herds, and enable the correct use of organic waste (stool and urine) from dairy activities (Embrapa Gado de Leite, 2020). This infrastructure has sensors to monitor the temperature, humidity of the environment, and lighting.

Provenance is defined as describing the origins of a piece of data and the process by which it arrived in a database (Buneman et al., 2001). The provenance of data can be used to aid in the authenticity, reliability, and reproducibility of decisions. From the provenance metadata, it is possible to form a critical basis for extracting this information for validation and information tracking. Provenance allows data to be tracked, shared, discovered, and reused, favoring collaborative activities. The standard model for specifying the provenance of data is the PROV (Belhajjame et al., 2013). Few articles discuss provenance in the agribusiness context. Da Cruz et al. (2019) discuss the use of data provenance in the agricultural context, considering problems such as the irreproducibility of experiments. The article also presents "Rflow", a framework for managing workflows. Da Cruz et al. (2018) present the OpenSoils program. OpenSoils architecture is driven by open provenance and aims to improve the reproducibility of experiments and deliver high-quality datasets, knowledge, and provenance-based maps. In these studies, the support for the decision-making process is not explored, nor is it used to mitigate past decisions. e-LivestockProv architecture was proposed to address aspects related to the use of data provenance to support decision-making.

3. e-Livestock Architecture

To support traceability in the decision-making context, e-LivestockProv architecture was designed to collect and process data generated by sensors attached to the animals

and add additional information external to the environment related to provenance. Through the capture of provenance, it is possible to obtain the source of the information that supported a decision and thus, improve the decision process. E-LivestockProv is composed of four layers, as shown in Figure 1.

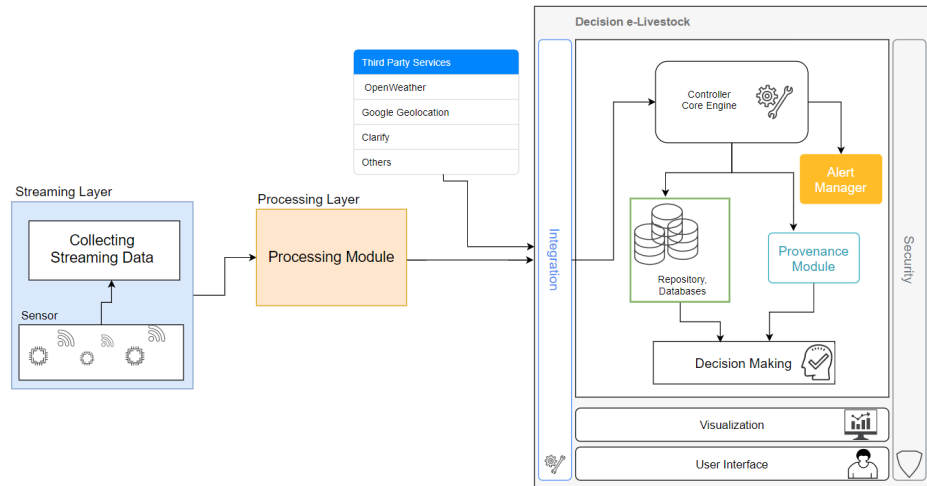


Figure 1. e-LivestockProv Architecture Overview

Streaming Layer: This layer is responsible for collecting data generated by sensors used in the physical space, such as devices connected to the animal's body or environment. This layer is designed to handle large volumes of data, ensuring availability and fault tolerance for the collected data.

Processing Layer: This layer is responsible for processing the collected data to be integrated with information from other sources, services, and external APIs, such as context data, environmental information (temperature, humidity, and weather forecast). The main advantage of this layer is supporting integration and interoperability.

Controller Layer: It is responsible for the business rules that will generate automated alerts in decision support, for example, if an animal's production is below standard. The controller layer also sends data to the repositories to be persisted and used by the visualization layer dashboards.

Visualization Layer: It allows the researcher to visualize the data in real-time, according to a time interval, through a panel. The researcher can also analyze and interpret data at different granularities.

3.1. Data Provenance

To capture the provenance of data, the PROV model was extended to accurately capture the provenance of PLF-related data. Figure 2 presents a partial view of the model. Entities represent animals, agents are researchers, and activities are actions performed on the smart farm. Activities can be described as insemination, milking, or processing the collected data. Through the provenance of these data, it is possible to identify the data sources and the interactions that researchers and users carry out, tracking decisions related to these specific activities.

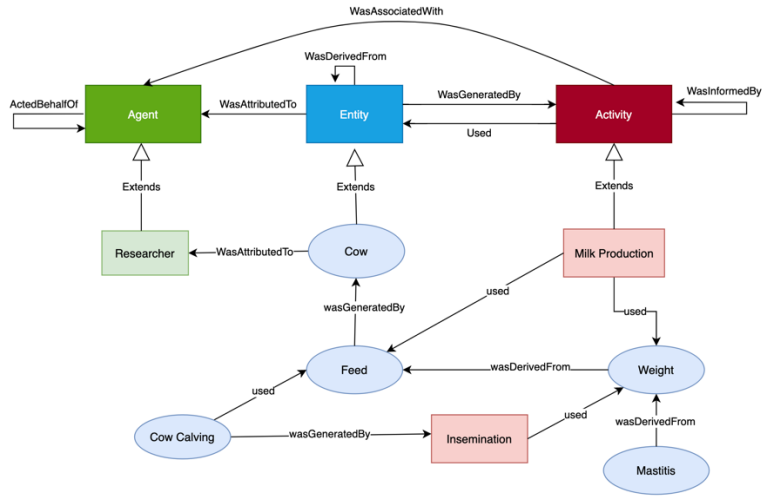


Figure 2. Provenance Model for e-Livestock

Considering the provenance model shown in Figure 2, it is possible to answer questions such as: (i) Did the production of an animal that was feeding correctly, with adequate weight, diminished due to the occurrence of inflammation (mastitis)?; (ii) Was the animal's weight loss due to a change in diet that the animal did not adapt to?; (iii) Was an animal's weight loss due to a parturition event?; (iv) Did the average production of an animal drop due to a temperature change?; Did the temperature variation cause the animal to spend more energy maintaining body temperature instead of producing milk?; (v) Did the increase in the average number of mastitis cases, due to an increase in humidity, favoring the proliferation of environmental bacteria?; among other questions.

3.2. Evaluation

To verify the feasibility of the proposed architecture, a Proof of Concept (PoC) (Yin, R. 2013) was specified. We defined its scope based on the GQM (Goal, Questions, Metrics): “To **analyze** the use of provenance captured by the architecture **from the point of view** of researchers for decision making, **in the context** of the Compost Barn production system with a focus on livestock”.

As a result, the following research question was defined: RQ. **How can provenance capture in e-LivestockProv architecture support the decision-making process on farms by tracking interactions and relationships between data?**

The evaluation was carried out based on collecting and processing data from the *Compost Barn* production system (Embrapa Gado de Leite, 2020), located at EMBRAPA - Gado de Leite, experimental field. This system is part of a research project with Brazilian and international institutions related to improving the dairy cattle production system. Cows are organized into lots according to their production and can change lots over time. Continuous monitoring allows for adjustments in the animals' living conditions, enabling increased production and quality of life.

In this PoC, we use an offline dataset obtained in the field. This dataset contains the records of weight (kg), milk production (L), and mastitis of a given animal, in addition to the averages of temperature (°C) and humidity (%) over the months in the Compost Barn. Table1 shows a partial view of this dataset, where the mastitis field

indicates the positive (true) or negative (-) incidence of the disease in the animal. Mastitis is an inflammation of the mammary gland that reduces milk production and causes discomfort to the animal. In addition, it generates costs with medication and technical assistance for the treatment.

Table 1. Animal Dataset

Month	September	October	November	December	January	February
Weight	650	624	630	664	671	683
Milk	-	-	38.8	46	45.1	46
Mastitis	True	True	-	-	-	-
Umidity	98	99	88	89	89	85
Temperature	26.1	28.1	24	24	22.4	24.8

As data is collected, processed, and made available on the dashboards, it was possible to monitor the Compost Barn environment. As example, we can mention: (i) the increased weight of animals, (ii) the increased or decreased milk production per batch, and (iv) the Compost Barn temperature and humidity variation. With the monitoring, we identified a reduction in the milk production of one of the lots. Through the provenance capture, it was possible to investigate the reason for discarding the milk. In September and October, there was a high humidity level in the Compost Barn (highlighted in red in Table 3), which generated a proliferation of environmental bacteria. These bacteria caused mastitis in the animal, which needed to be medicated and lost weight. Thus, due to the medications, the decision was made to discard the milk. Another case was detecting the sudden increase in weight of one of the cows and its change of lot. By tracking the evolution of animal weights, a peak was observed in one of the lots. With the provenance model, we identified that one of the cows in that batch had an insemination event, causing an increase in weight and later deciding to migrate a batch. In this way, the e-LivestockProv architecture provided decision support through traceability and the combination of sensor data with context data. Data are displayed in graphs and analyzed by the researchers. Therefore, considering our research question (RQ), once a decision or event considered unusual was detected, it was possible to clarify the reason and the entire process that generated the decision through the relationships captured by the provenance model.

4. Final Remarks

Decisions taken in agriculture need to be traceable due to the diversity of data and devices present in different contexts. This work presents an architecture supported by a provenance model for real-time data collection, processing, and visualization to support decision-making. By capturing the origin of the data and the processing steps, we can investigate events and track decisions made in the environment. Through an initial evaluation, it was possible to verify the feasibility of the architecture considering the relationships observed in the provenance model applied in real agricultural environments.

As future work, we intend to integrate other data sources such as social networks and evolve the provenance model. In addition, it is necessary to generate other instances of e-LivestockProv architecture and associate them with a software ecosystem and explore aspects of collaboration, communication, and integration between farms to support decisions in these instances. Furthermore, combining and processing additional data sources and sensors can lead to more accurate results, reduce costs and maintain agribusiness sustainability.

Acknowledgments

We would like to thank the researchers who participated in the evaluation and express thanks to the National Research Center of Dairy Cattle (Embrapa Dairy Cattle) of Brazilian Agricultural Research Corporation (Embrapa). This work was partially funded by UFJF/Brazil, CNPq/Brazil (grant 311595/2019-7), and FAPEMIG/Brazil (grant APQ-02685-17), (grant APQ-02194-18). This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

- Bahlo, C., Dahlhausac P., Thompsonac, H. and Trotterbc, M. (2019) “The role of interoperable data standards in precision livestock farming in extensive livestock systems: A review”. *Comp. and Elect. Agric.*, p. 459-466.
- Banhazi, T.M., Lehr, H., Black, J.L., Crabtree, H., Schofield, P., Tscharke, M., Berckmans, D. (2012) “Precision livestock farming: an international review of scientific and commercial aspects”. *Int. J. Agri. Bio. Eng.* 5 (3), p. 1–9.
- Belhajjame, K., Cheney, J., Coppens, S., Cresswell, S., Gil, Y., Groth, P., Klyne, G., Lebo, T., McCusker, J., Miles, S., Myers J., Sahoo, S., Tilmes, C., Moreau, L., and Missier, P. (2013) “PROV-DM: The PROV Data Model”, W3C Recommendation REC-prov-dm-20130430, World Wide Web Consortium.
- Buneman, P., Khanna, S. and Tan, C. (2001) “Why and where: A characterization of data provenance”, 8th Int. Conf. on Database Theory, p. 4-6.
- Da Cruz, S. M. S., Ceddia, M. B., Miranda, R. C. T., Rizzo, G., Klinger, F., Cerceau, R., Cruz, P. V. (2018) “Data Provenance in Agriculture”. *Int. Provenance and Annotation Workshop*. Springer, Cham, p. 257-261.
- Da Cruz, S.M.S., Do Nascimento, J.A.P. (2019) “Towards integration of data-driven agronomic experiments with data provenance”. *Comp. and Elect. Agric.*, 161, p. 14-28.
- Embrapa Gado de Leite. Brasil tem a primeira instalação de compost barn destinada à pesquisa - 2020. Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/53360675/brasil-tem-a-primeira-instalacao-de-compost-barn-destinada-a-pesquisa>>. Accessed in 03 mar 2021.
- Gomes, J., David, J. M. N., Braga, R., Ströele, V., Arbex, W., Barbosa, B., Gomes, W., Fonseca, L. (2021) “Architecture for Decision Support in Precision Livestock Farming”. *Proceedings of the 15th Brazilian e-Science Workshop*. SBC, p. 41-48.
- Karthick, G. S., Sridhar, M., & Pankajavalli, P. B. (2020) “Internet of things in animal healthcare (IoTAH): review of recent advancements in architecture, sensing technologies and real-time monitoring”. *SN Computer Science*, 1(5), p. 1-16.
- Villa-Henriksen A., Edwards G., Pesonenc, L., Greenbd O. and Sørensen, C. (2020), “Internet of things in arable farming: Implementation, applications, challenges and potential”. *Biosystems Eng.*, p. 60-84.
- Yin, R. (2013). *Case study research: Design and methods* Sage Publications, Inc; 5th ed.
- Zhai, Z., Martínez, J.F., Beltran, V., Martínez, N.L. (2020) “Decision support systems for agriculture 4.0: Survey and challenges”. *Comp. and Elect. Agric.*, 170, p. 105256.