# 434. The infrastructure for cattle data at the Swedish university of agricultural sciences, Gigacow

T. Klingström[*], I. Ohlsson[1] and D.J. de Koning

*Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 7023, 750 07 Uppsala, Sweden; tomas.klingstrom@slu.se*

## Abstract

Research in animal breeding and precision livestock farming both require vast volumes of highly heterogeneous data, often beyond the capabilities of the universities' own research farms. To facilitate access to detailed phenotypes from a network of farms, we established the infrastructure for cattle data at the Swedish University of Agricultural Sciences (SLU Gigacow). A main achievement of SLU Gigacow has been the development and implementation of automated daily collection of records from each participating farm's herd management system to a centralized server at SLU. SLU Gigacow will monitor around 5,000 cows (including annual replacements) when it is fully operational. Although still under development, SLU Gigacow has already received much attention from various stakeholders, as well as media. In the long-term we expect SLU Gigacow to become a valuable infrastructure for discoveries, as well as a testbed for implementation of new technologies.

## Introduction

While a university can do much research in their own dedicated research farms, many studies, especially those related to genetics and genomics, require much larger sample sizes and access to field data. It is very demanding and highly inefficient for individual researchers to recruit farms that are willing to contribute data for each individual research project. To address this challenge, the infrastructure for dairy cattle data at the Swedish University of Agricultural Sciences, Gigacow (SLU Gigacow), was founded in 2018. Its main purpose is to support research projects in dairy cattle genetics, precision agriculture, and other research concerning dairy farms and the environment. In a modern dairy, farm machines and sensors automatically collect data on milk composition, activity, and feeding behaviour of cows to ensure good animal health and high productivity. Genotyping of calves for estimation of genomic breeding values is now also routine procedure at many farms and genotyping with a 45K SNP-chip is available as a commercial service at a low cost. At a farm, the pedigree of each cow is known and the herd structure ensures that between 60 and several hundred animals are exposed to the same environment.

The purpose of SLU Gigacow is to establish a single data collection platform useful to researchers at SLU and elsewhere to accelerate the scientific process. Once an infrastructure has been established, large amounts of data can be collected at a low cost. The cost of establishing a data collection infrastructure with legal agreements and software solutions for automated data collection and data curation is however a major investment. The limitations of funding and a lack of dedicated infrastructure have been identified as the major bottlenecks in phenome research in non-model organisms, as projects are either narrowly focused on a single biological pathway in response to phenotypic change, while ignoring the rest of the dataset, or limited to broad sweeping conclusions about the dataset based on shallow interpretations (Burnett *et al.*, 2020). By providing a rich dataset of phenotypes and associated data, researchers can better plan the scope of projects and engage in exploratory projects with minimal overhead and startup costs. Thus, SLU Gigacow can provide the mechanism necessary to encourage junior and senior investigators to conduct focused, hypothesis-driven research framed in a genome to phenome context and to critically consider experimental design, power, and sample sizes when launching '-omics'-based experiments.

## Materials & methods

Production data from farms is collected by the software Gigacow Kollector, written in Python 3 (Beaverton, USA), and automates export of data from herd management systems such as DeLaval DelPro (Tumba, Sweden) and Lely T4C (Maassluis, the Netherlands). Data in certain reports on the DeLaval DelPro can be scheduled for automatic exports and other sensor manufacturers also provide similar solutions, enabling data from the herd management system and other equipment to be automatically exported to the temporary file storage area created by Gigacow Kollector. Each day, Gigacow Kollector runs a scheduled task to upload data with Secure File Transfer Protocol (SFTP) to a file storage server located at SLU, which stores the raw files. A verification script run on the server side ensures that all expected reports are uploaded. Data from the SLU Gigacow data collection server is then imported to the Gigacow SQL Server database (DB) by means of multiple processes for extraction, transformation and loading that are maintained with SQL Server Integration Services (SSIS) (Microsoft, Redmond, USA). Researchers can download data directly from the DB via a database data connection or, to facilitate analyses, by connecting the tool of their choice, such as Excel, to a regularly preprocessed and aggregated Online Analytical Processing cube deployed with SQL Server Analysis Services (SSAS).

In addition to data collected on farms, SLU Gigacow also collects data through an agreement with Växa Sverige AB (Uppsala, Sweden), which is a service that aggregates information on subjects such as claw health, veterinary visits, and other key performance indicators related to animal health and productivity. This data is compiled quarterly in fixed length files and sent to the SLU Gigacow server using SFTP. Genome analysis of animals is performed by the service Växa Genomanalys using the 45k EuroG MD beadchip (Amsterdam, the Netherlands). Calves born in herds that participate in SLU Gigacow are genotyped as a part of the tagging process, in which a sample is collected by the farmer and submitted to the genome analysis service. Adult animals are sampled by the farmers or visiting technicians from Växa by collecting blood or hair samples from the tail. Ethics approval from the authorities is required and has been granted for collection of blood samples from adult animals but is not required for hair sample collection.

The data collection system of SLU Gigacow is under continuous development but, in its current form, presents a modular solution to quickly integrate information from many different sources (Figure 1).

## Results

Data collection from dairy farms was delayed by the Covid-19 pandemic but was started by remote installations in 2021. In a steady state situation, approximately 5,000 animals on 15 farms will be monitored on an annual basis, with a budgeted annual replacement rate of 35%. So far 1300 animals have been genotyped in collaboration with Växa and the farmers. Using Växa Sverige's commercial service for genome
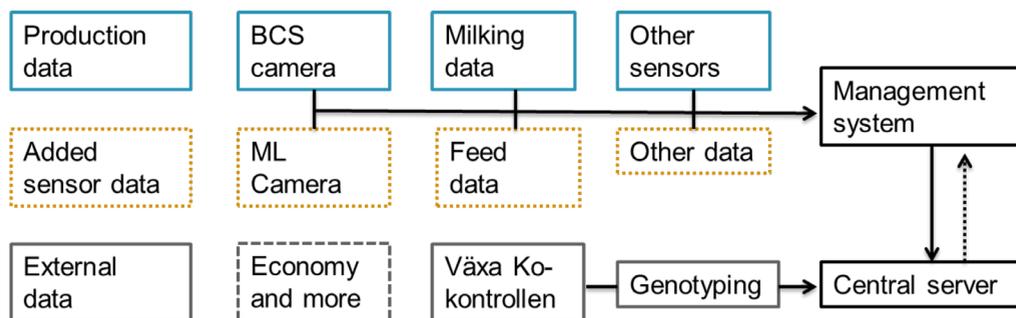


**Figure 1.** Data collection in Gigacow. Dotted boxes are currently under development.

analysis, the cost per genotyped calf is ~25 Euro at an exchange rate of 0.1 SEK per Euro. Genotyping of adult animals is done by arrangement and cost between 32 and 36 Euro, depending on the size of the farm, who performs the sampling, and labour required to collect animals.

A number of key performance indicators have been tracked during the establishment of SLU Gigacow (Table 1). The SLU Gigacow infrastructure was established not only to support genomics research but also to serve as an interdisciplinary infrastructure that serves all fields related to dairy farming. Education of agronomists specialising in agricultural technology was abolished in Sweden ~20 years ago and SLU is currently working to increase its focus on agricultural technology, bioinformatics, and precision farming. Establishment of SLU Gigacow has created a natural contact point at SLU for stakeholders in the industry and at other universities to bring technology developers and the farming industry closer together. SLU Gigacow has therefore been used as an example in multiple articles and conferences on smart farming and data collection. SLU Gigacow has also participated in numerous grant applications with partners at Linköping University, Uppsala University, Dalarnas University, the University of Oslo, and several nodes in the Elixir European life science infrastructure for biological information (ELIXIR). The total approval rate for grant applications supported by SLU Gigacow has so far been 42%, compared to an average funding rate of 14% for the Swedish Research Council for Sustainable Development (FORMAS) and 20% for grant applications in Natural science and technology by the Swedish Research Council (VR).

Technology development and testing is also an important topic, as the implementation of smart solutions is a key area in the digital transformation of the agricultural sector (OECD, 2019). Gigacow is developed to function as a testbed for innovation and is a member of the Nordic Testbed Network (https://nordictestbednetwork.se/gigacow-2/). Data from farms is continuously collected, facilitating both longitudinal (before and after implementation) and case-control studies, enabling users to 'test before invest'. Activities in this role include projects integrating Edge Artificial Intelligence into farming decision support systems, in collaboration with Uppsala University and a major equipment manufacturer, and assessment of market acceptance for remote sensing applications, together with a major producer of visible, near infrared, and short-wave infrared sensors. To formalise this role, SLU Gigacow also participates in the application for a European Digital Innovation Hub, in collaboration with Linköping University, the Research Institutes of Sweden, and multiple regional actors for technology and innovation.

## Discussion

The major challenges during the development of SLU Gigacow have not been how to collect data but to identify what data there is to collect, what agreements are necessary with various stakeholders to collect the data, and finding a robust technical infrastructure that enables the collection of data from heterogeneous

**Table 1.** Key performance metrics tracked in SLU Gigacow.

| Indicator | Number |
|---|---|
| Animals genotyped | 1,294 |
| Articles and presentations of Gigacow | 19 |
| Projects supported with data collection | 3 |
| Projects supported with expertise | 5 |
| Supported applications | 20 |
| Supported applications funded | 8 |
| Supported applications not funded | 11 |
| Supported applications pending | 1 |

data sources and present it in an accessible manner. There is fierce competition in the agricultural sector, both between equipment manufacturers and service providers, which makes protection of international property rights a key concern for these commercial companies. At the same time, partnering with the industry can yield considerable benefits of scale, as well as new research opportunities. The estimated cost of genotyping at the National Genomics Infrastructure would, despite an academic subsidy have amounted to ~52 Euro per animal, even before adding costs for sample collection and DNA extraction, which would have had to be performed in-house at SLU. There is also a strong interest, both among equipment manufacturers and service providers, to further increase the precision and scope of data collection in farming. Close collaboration between researchersm farmers and other stakeholders in primary production can thereby both yield new technology for the study of animals and insights that motivate initiation of new research.

## Conclusions

Digital transformation is rapidly changing all sectors of society and many sectors face a lack of computer scientists and bioinformaticians who can process and analyse biological data (Batut *et al.*, 2018). Data collection and analysis is, therefore, an increasingly large part of the workload for many researchers (Klingström, 2017). Automation and collaboration with industry is, therefore, important to reduce the workload of researchers by making sure that data is readily available in a format that is accessible to both domain experts and technology developers who are aiming to bring technological innovations into new domains. Thus, the establishment of livestock data infrastructures can accelerate genetics research by genotyping and phenotyping animals at a scale that is not feasible outside the agricultural sector.

## Acknowledgements

## References

Batut B., Hiltemann S., Bagnacani A., Baker D., Bhardwaj et. al (2018). Cell Systems 6, 752-758.e1. https://doi.org/10.1016/j.cels.2018.05.012

Burnett K.G., Durica D.S., Mykles D.L., Stillman J.H., Schmidt C., (2020). Integrative and Comparative Biology 60, 397–401. https://doi.org/10.1093/icb/icaa059

Klingström, T., (2017). Data integration and handling, Acta Universitatis Agriculturae Sueciae. 1652-6880; 2017:99 Sveriges lantbruksuniversitet., Uppsala.

OECD, (2019). Digital Opportunities for Better Agricultural Policies. OECD. https://doi.org/10.1787/571a0812-en