A Gigacow example, weather data in longitudinal studies on heat stress

T. Klingstrom[1], J.I. Ohlsson[1], P. Ask-Gullstrand[1], Lena-Mari Tamminen[2] and D.J. de Koning[1]

[1]Department of Animal Biosciences, Swedish University of Agricultural Sciences, Box 7023, 750 07, Uppsala, Sweden
[2]Department of Clinical Sciences, Swedish University of Agricultural Sciences, Box 7054, 750 07, Uppsala
Corresponding Author: tomas.klingstrom@slu.se

Abstract

Large and varied datasets from modern precision livestock farming equipment can be used to tackle increasingly complex research questions such as defining a phenotype of heat stress resistance in dairy cattle milk production. Heat stress causes drops in production and changes in behavior that are difficult to detect unless the cow in question is monitored over longer periods and more variable conditions than conventional experimental designs may allow. We present here how the SLU Infrastructure for dairy data collection, Gigacow support complex data-driven dairy research using heat stress as a case study. SLU Gigacow gathers daily data updates from a set of Swedish commercial dairy farms with a digital farm management system (FMS) overseeing either a robot milking system or milking parlor, and links the data per-cow to individual 50k SNP genotypes and national animal database information including trade history, pedigree and health events. The way SLU Gigacow support the project *From Sensitive to Robust Athlete – Exploring the Opportunities of Genomic Selection to Help Dairy Cows Cope With Increasing Temperatures* provides a good case study on how the data and knowledge generated by:

- Validating an external dataset from the Swedish Meteorological and Hydrological Institute.
- Collecting daily milking data from farms over multiple lactations.
- Providing cost coverage to Gigacow farms for genome analysis services and collect genetic data generated from the genome analysis in collaboration with Växa Sverige and Nordic Cattle Genetic Evaluation.
- Continuously evaluating, and including new data analysis models to support researchers and ensure that lessons learned in different projects can be included in new studies.

The collection of high quality longitudinal datasets was a key motivation for the establishment of SLU Gigacow and the infrastructure actively tries to prioritise data collection efforts to support future research and industry needs.

Keywords: data collection, infrastructure, heat stress.
Presented at the ICAR Anual Conference 2024 in Bled at the Session 10: New approaches in the field of functional traits for management and breeding.

Introduction

As the digital transformation of society progress it is apparent that agricultural research institutions must adapt and ensure that a mixture of competency in data science and agriculture is maintained within the organization. The rapid pace of development also makes it difficult for researchers to keep up to date on new technology or data sources becoming available. Researchers affiliated with a data collection infrastructure can therefore have an important role in evaluating technology and ensuring that colleagues have access to new technologies to generate data for their research.

The SLU Infrastructure for dairy data collection, Gigacow (SLU Gigacow) is a data collection infrastructure at the Swedish University for Agricultural Sciences previously presented at ICAR 2023 (Ohlsson et al. 2023). The infrastructure collect data from a wide variety of sources and its role in the project *From Sensitive to Robust AthleteExploring the Opportunities of Genomic Selection to Help Dairy Cows Cope With Increasing Temperatures* provide an example of how a data collection infrastructure can support dairy research.

A SWOT analysis of SLU Gigacow

Working with a data-driven approach to develop a research infrastructure requires both patience and trust as the start-up process of a general data platform require more time and consideration than ad-hoc data collection for a single project. Looking at the Strengths, Weaknesses, Opportunities and Threats to the Gigacow infrastructure highlights the following table:

Strengths	Opportunities
Reduced technical overhead in projects	Integration of new data sources
Continuous improvement of data collection	Standardised APIs make big data accessible
Provides a platform for development	Support research on data models
Can integrate results in multiple projects	Systematic gap-analysis for new technologies
Threats	**Weaknesses**
Slow start before first results are generated	Prioritisation of data sources
Little funding for indefinite projects	Lack of data specialists in agriculture
Risk of obsolescence	False negatives are hard to detect

Timeline of the project and SLU Gigacow

The project application was written in the early spring of 2022 and funding began on 1 January 2023, data collection from the project however began already in 2020 as climate change and warming was identified as an area where researchers at SLU where likely to wish to use SLU Gigacow. In Sweden (Figure 1)

In this project data from the herd management system of farms participating in the SLU Gigacow network is combined with genotypes from the Nordic Cattle Genetic Evaluation and meteorological analysis data collected from the open data collection of the Swedish Meteorological and Hydrological Institute (SMHI).

To evaluate the need for on-farm temperature sensors SLU Gigacow has cross-referenced the Mesoscale Analysis (MESAN) model from the Swedish Meteorological and Hydrological Institute (SMHI) with historical measurements from SLU Lantmet, which collates climate data from weather stations distributed across the country. The MESAN model interpolates climate data from SMHI's national network of weather stations to model conditions across Sweden on a 11x11 km grid. Measurements from MESAN on temperature and humidity were deemed to be of

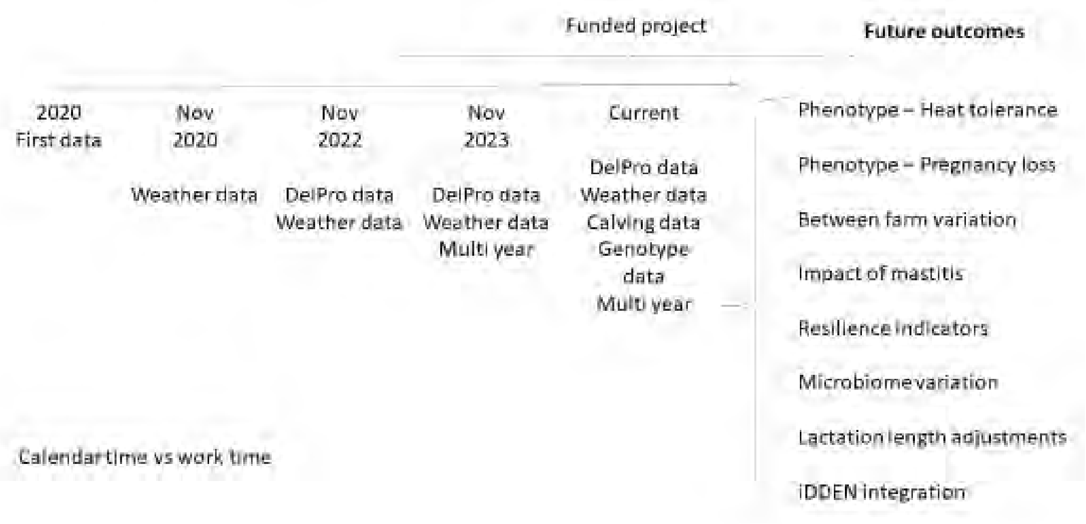

Figure 1. Timeline of the project and SLU Gigacow.

sufficient quality while wind and precipitation could differ significantly when comparing on-site measurements with the MESAN model. MESAN model temperatures from each farm's matching grid square was therefore used to infer the Temperature-Humidity index at each farm in the project.

Strengths

- Faster per project
- Continuous improvements to data collection
- Provides a platform for development
- Provide an integration platform.

Opportunities

- Integration of new data sources
- API:s "only" legal issues limit data.
- Methods development
- Gap analysis
- Model building.

Threats

- Slow start
- Long term funding
- Obsolescence

Weaknesses

- Prioritisation
- Recruiting
- Missing data
- Dedicated staff
- Data separation in "data lakes"