

**FARM-TO-TABLE SCIENCE: DAIRY DATA MINING FOR FUTURE RESILIENCE**

Ingemar Ohlsson, Tomas Klingström.

Swedish University of Agricultural Sciences, Uppsala, Sweden.

Data in Farm Management Systems (FMS) present a wealth of information from a large number of sensors and recordings but are designed to provide a current overview, discarding older information that is vital to long-term studies of a herd's genetics, behaviour, and environment. Most farms with automated milking systems (AMS) in the Nordic region have not yet invested in FMS configured to exchange data according to the Animal Data Exchange (ICAR ADE) or participate in the International Dairy Data Exchange Network, meaning that the data is discarded unless measures are taken to collect the data.

In this work, we present the infrastructure for dairy cattle data at the Swedish University of Agricultural Sciences, Gigacow (SLU Gigacow), which collects data from a set of Swedish dairy farms, all of which use some form of AMS. Each farm's FMS sends nightly reports to SLU Gigacow, where records are harmonized and collected in a central database. Collected records include milking statistics, health events, traffic data, and SNP genotypes for thousands of cows, and are made accessible to researchers through SQL queries with an optional R wrapper. SLU Gigacow also integrates data from the Swedish national cattle registry, including pedigrees and herd transfers for cows resident at participating farms.

SLU Gigacow's longitudinal observations (first data collected in 2019) link genotype to phenotype and animal welfare with the goal of accelerating pilot studies in dairy science, as well as providing a big dataset from cows in active, commercial settings. The data collection software written in Python 3 currently collects from DeLaval DelPro, and can be extended to any FMS with a graphical user interface running on most consumer operating systems. After harmonization to resolve differences in language and FMS versions, data are stored in a database maintained at SLU with SQL Server Integration Services (SSIS). By agreement with Växa Sverige AB, participating farmers also get a large number of animals genotyped using the 45k EuroG MD beadchip. Currently, the database includes information on over 17,000 cattle, over 3,000,000 milkings, and 2,969 SNP genotypes. The cross-referenced data can be mined for various purposes, including stress responses and resilience traits.

While SLU Gigacow is intended to collect from Swedish farms and support Swedish researchers, it serves as a proof-of-concept that data from diverse sources and systems at dairy farms can be automatically gathered and collated in a researcher-friendly format. We believe that this shows the great utility of farm-to-table statistics and increased FMS interoperability. SLU Gigacow was constructed essentially without standardized interfaces for dairy data communication.

Establishment of such data standards would enable integrated measurements of even more facets of the dairy farm environment, create new niches for precision livestock farming equipment, and open great new vistas of information to drive adaptation to changing climates.