A Central Database for the Australian Dairy Industry

S. Jenkins[1], *T. Francis*[1]

[1]*DataGene Ltd., AgriBio, 5 Ring Road, 3083 Bundoora, Victoria, Australia*

Abstract

In 2012 the Australian Dairy Herd Improvement Scheme began work on developing a replacement for their outdated genetic evaluation system. This project was entitled GESII (Genetics Evaluation System II). In late 2015 the industry looked to create a single industry body in Australian to manage genetic evaluation, data and milk recording standards and training. This along with other synergies enabled the GESII development to merge with existing milk recording software by the way of creating a Central Data Repository (CDR). This new single project is entitled GESNP (Genetics Evaluation System New Platform) and is delivered by DataGene, the newly established industry body. The goal is to provide a central industry data store where previously disparate and unconnected data can be shared and analysed.

DataGene established, with the broader industry a 'Data Governance Group' to manage the rules, standards and data management guidelines for data entering CDR.

The GESNP program currently has a technical delivery team of 4 business analysts and 25 software engineers. There are 6 subject matter experts from across DataGene and Department of Economic Development, Jobs, Transport and Resources (DEDJTR)

The first stage of the GESNP platform delivery will include the replacement of the legacy Genetic evaluation system and phase I of CDR. This will be delivered for beta and internal release in June 2017, with further phases staged for release into 2018.

CDR will feature an open architecture, and the data will be a combined dataset of all industry data without duplication and with validation, verification and standards agreed and applied. The Data Governance Group (DGG) will be instrumental in setting these agreed parameters. There will however, be conditions of entry to CDR which will include committing to the terms of a Service Level Agreement (SLA). Software development partner will be encouraged to innovate and deliver new services on this combined data for the greater industry good.

The CDR component of the GESNP program is broken into a phased approach with Phase I 'Acquire the Data' featuring data collection, governance and support for operational systems. Phase II known as 'Report and Inform' includes data analytics, increased functionality in reporting and analysis to better drive business decision making. Additional data providers will also come on stream during this phase. The final program phase, Phase III, 'Enhance and Extend' sees increased analysis using 'Big Data', use of IoT, AI and machine learning, identifying new areas of complimentary data such as 'Social media' and international data that can improve our understanding of customers, behaviours and trends. This phase will help provide 'predictive analytics' and 'what if scenario planning'.

It is expected that CDR will then enter a product lifecycle development phase of continuous improvement and enhancement to always stay relevant and valuable to the user community and our industry partners.

Keywords: central data repository, genetic evaluation, software, Australia