Final report

Data Quality Metrics for Sheep Genetics and BREEDPLAN

Project code: L.GEN.2004

Prepared by: Sarita Guy and Daniel Brown
                    Animal Genetics and Breeding Unit, University of New England

Date published: June 2021

PUBLISHED BY
Meat & Livestock Australia Limited
PO Box 1961
NORTH SYDNEY NSW 2059

Abstract

Data is the cornerstone of our genetic evaluation systems. Maximum data quality enhances accuracy of estimated breeding values (EBVs) and increases ability to make accurate selection decisions, which drive genetic gains. This project developed metrics to describe the quality of data submitted for inclusion in genetic evaluation systems. The work focused mainly on the sheep genetic evaluation systems delivered through Sheep Genetics, and the refinement and reporting of the data quality metrics currently reported in the RAMPing Up Genetic Gains (RUGG) reports.

This project achieved 4 key outcomes:

1. **A data quality framework** to characterise the quantity and quality of genetic evaluation data for Terminal, Maternal and Merino flocks. This prototype framework included:
    *   14 new and 7 existing data quantity, quality and timeliness-related metrics.
    *   A Data Quality Score (DQS) and star rating to characterise the overall quality of each flock's data.
2. **Demonstration of value proposition**
    The DQS was developed independently of, but is related to, metrics that describe genetic gain. Flocks with higher scores had higher index accuracies and greater rates of genetic progress.
3. **Implementation of the data quality framework**
    OVIS software has been updated to calculate new metrics, DQS, and automatic identification of data recording strengths. and recommendations for improved data recording.
    Software has been developed to generate interim DQS reports to demonstrate how it can be incorporated into RUGG reports. These changes are anticipated to be implemented in the next phase of RUGG report development.
4. **Successful road-testing** in collaboration with Sheep Genetics.
    The data quality framework and DQS was road-tested at 6 events involving 96 flocks. The DQS was well received by service providers and sheep breeders, and there was overwhelming support to publish the DQS star ratings for each flock after a grace period.

The data quality framework will assist ram breeders to target improvements in their recording programs. On a wider industry level, the framework will help identify and highlight breeders who collect and submit high quality data, and provide increased transparency to ram buyers about the quality of data used to calculate EBVs. There is also potential to further expand the framework to the valuing of phenotypes and the extent and/or value of each flock's data contribution to the reference population.

This proposed data quality framework requires continuous monitoring and refinement of weighting factors over time. Appropriate tools and support are also required to help breeders understand how the information generated through this framework can be used to improve their data recording strategies. Recommended future research and extension includes: studies of the cost-benefit of recording better quality data; better demonstration of data quality and genetic gain outcomes; and further development of a framework to value data contribution to reference populations. The demand for an updated data quality framework for BREEDPLAN (currently delivered through DataAudit software) is unknown. In principle, frameworks for evaluating data quality should be consistent across species, primarily to facilitate extension, and potentially to simplify introduction of systems for valuing data for reference populations.

Executive summary

BACKGROUND

Data is the cornerstone of our genetic evaluation systems. Maximum data quality enhances accuracy of estimated breeding values (EBVs) and increases the accuracy of selection decisions, which in turn drives higher rates of genetic progress. The genetic evaluation systems for the Australian red meat industry, Sheep Genetics and BREEDPLAN, primarily rely on industry data submitted by seedstock breeders. While there are standards and guidelines for data submission, there is wide variation in data recording and submission across the industry.

This project developed metrics to describe data quality used in genetic evaluation systems. The work focused mainly on the sheep genetic evaluation systems delivered through Sheep Genetics, and the refinement and reporting of the data quality metrics currently reported in RAMPing Up Genetic Gains (RUGG) reports, which are provided to breeders by Sheep Genetics.

OBJECTIVES

All project objectives were achieved for the Sheep Genetics evaluation system:

- **Objective 1. Demonstration of the value proposition** of data quality metrics in relation to prediction of genetic merit

- **Objectives 2 & 3. Development of applicable data quality metrics and refinement of the current data quality reports and metrics** within Sheep Genetics, as well as any additional metrics from BREEDPLAN

- **Objective 4. Demonstration of ways to calculate and report data quality metrics** on a per flock basis to the public

METHODOLOGY

- **Objective 1. The value proposition**

The value proposition for the current data quality metrics provided in RUGG reports was examined for Terminal, Maternal and Merino flocks. This was achieved by analyzing the relationships between the data quality metrics and 1) genetic gains, 2) regression of progeny performance on sire EBVs, 3) genetic parameter estimates.

- **Objectives 2 & 3. Development and refinement of the data quality metrics**

A review of existing frameworks for characterization of genetic evaluation data was conducted. A data quality framework was developed with 3 components – quantity, quality and timeliness.

Current metrics reported in RUGG reports were categorized as data quantity or data quality characteristics. Current metrics were refined, and additional metrics developed and tested.

- **Objective 4. Demonstration of reporting**

A Data Quality Score (DQS) was derived to characterise the overall quality of each flock's data. Three different methods were explored to decide which metrics to use, and their importance in the score (i.e. weightings):

  o Principal component analysis (PCA) of all metrics

- o Stepwise multiple linear regression (MLR) trained on index accuracy, average index and index trend, and the first principal component of all 3 gains metrics
- o Informed weightings: a weighting structure based on results from exploratory analysis of metrics, PCA and MLR; combined with industry knowledge and experience.

The AGBU software behind the RUGG reports and the Sheep Genetics website were both updated to include the new data quality metrics. New software was developed to generate interim DQS reports on an individual flock level. In collaboration with Sheep Genetics, this reporting was road-tested in workshops with service providers and breeders. Individual flock RUGG and DQS reports were provided, and workshopped to obtain feedback.

RESULTS/KEY FINDINGS

- **Objective 1. The value proposition**

Most current data quality metrics were significant predictors of genetic gains, explaining between 2 to 60% of the observed variation in the rate of progress between flocks. Flocks with higher quality data made more genetic progress. There was small but statistically significant relationships between data quality metrics and sire-progeny regression and genetic parameter estimates. However, these results were variable between breeds and the current data quality metrics only explained a small proportion (on average ~5%) of variation observed. Nevertheless, the wide variation in the quality of data submitted demonstrates opportunities for improved data recording strategies, and also highlights the need for better metrics describing data quality.

- **Objectives 2 & 3. Development and refinement of data quality metrics**

The current RUGG data quality metrics were refined, and additional metrics (including DataAudit-inspired metrics) were calculated. There were 4 quantity, 9 quality and 5 timeliness-related new metrics calculated. Some metrics explored (e.g. chi-squared statistic for digit frequencies, maximum frequencies of single values) may have limited value as feedback to breeders as frequencies did not deviate from what was theoretically expected. Nevertheless, these metrics can still be used as a way to highlight any suspicious data.

- **Objective 2 & 3. Development of an overall Data Quality Score (DQS)**

To combine the quantity, quality and timeliness metrics, an overall Data Quality Score (DQS) for each flock was derived using 3 methods. Detailed investigation of the alternate methods yielded no definitive or optimal method for how an overall DQS is calculated. All resulting DQSs were related to genetic gains, and the DQS were also moderately to highly-correlated with each other.

The informed weightings approach to derive the overall DQS is recommended, which utilises 21 metrics. The calculation of this DQS did not include metrics describing the amount of genetic gains achieved by each flock, but they were moderately to strongly correlated. This approach also provides a balance between scientific rigour with industry adoption since it is easily understood and interpreted.

- **Objective 4. Reporting of data quality metrics and score**

This project proposes the incorporation of the DQS and star ratings into Sheep Genetics RUGG reports:

- DQS be scaled from 1 to 99 for ease of interpretation

- Categorised DQS into 'star ratings' through equal ranges, with star ratings allocated for as 1 star: 0-20, 2 star: 20-40, 3 star: 40-60, 4 star: 60-80 and 5 star: 80-100
- Data recording strengths and recommended areas for improvement highlighted for each flock

OVIS software has been updated to calculate new metrics, DQS score, and automatic identification of data recording strengths and recommendations for improved data recording. New software was developed to generate interim DQS reports to demonstrate how it can be incorporated into RAMping Up Genetic Gains reports. Below is a screenshot of page 1 of the DQS report.

RAMping Up Genetic Gain

MOCK UP report to include Data Quality Score prototype

RAMping Up Genetic Gain reports aim to provide breeders with better information on the variables that impact the rate of genetic gain of their flock. These variables include the amount and quality of pedigree, the quality of the data and how effectively this data is used. By considering the individual components of the breeders equation and balancing them breeders are able to maximise genetic gain.

The RAMping Up Genetic Gain reports help evaluate your flock's breeders equation. RAMping Up Genetic Gain reports are a tool breeders can use to summarise data from their flock into these components that influence genetic gain.

The context information supplied below each report section provides a benchmark of how your flock is performing relative to the rest of the analysis for factors impacting genetic gain. Context information includes:

- Flock Average - the five year average of the flock the report is based on for the component in question
- Analysis Average - the five year analysis average for the component in question
- Rating - a five-star rating of how your flock compares to the rest of the analysis for that component of genetic gain. For example a five-star rating means that your flock is in the top 20% of the analysis for that component. A one-star rating means your flock is in the bottom 80% for that component.
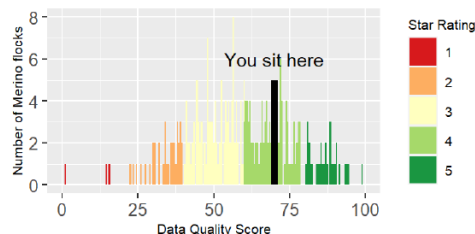
Flock Code: GENERIC. **Site Code:** GENERIC (MERI flock). **Flock Name:** GENERIC

Date Created: 31 May 2021 from run 16May21.

Data Quality Score: 69.9

★★★★

4 star

Your Data Quality Score (DQS) is an overall indicator of the quantity, quality and timeliness of your data.

Your data recording strengths:

- Lack of syndicate pedigree
- Recording of weight traits
- Recording of wool traits

To improve your data quality:

- Submit pedigree data earlier
- Increase effective progeny numbers
- Submit wool trait data earlier

Although key recommendations have been provided, refer to the entire RUGG report to decide which components of data recording are most relevant to your flock.

- **Objective 4. Road-testing of DQS with industry**

The DQS prototype was road-tested at 6 events involving 96 flocks. The prototype was well-received and constructive feedback was provided to further enhance the usefulness of the DQS. There was overwhelming support to publically publish the data quality star ratings for each flock after a grace period.

BENEFITS TO INDUSTRY

- **An enhanced data feedback tool for breeders.**
  The RAMping Up Genetic Gains reports can be further enhanced with new data quality metrics, Data Quality Score, star rating, recommendations and strengths. This provides targeted advice to breeders to assist in management changes, improve data collection and submission and hence ASBV accuracy. In turn, this will assist in more accurate selection decisions and increased rates of genetic progress
- **Transparency for ram buyers** about the quality of data used to calculate EBVs. While EBV accuracies are available for individual rams, a DQS provides an indication of the overall quality of the flock's data and allows direct comparison across flocks.
- **A way to identify and highlight breeders** who collect high quality data. This could be used as a basis for discounted registration fees, or through breeder awards, or other signals and/or rewards
- **Engagement tool** for Sheep Genetics development officers (and service providers) for targeted extension activities for flocks with poor data quality.
- This data quality framework can be further developed to determine and value **data contribution to reference populations**

FUTURE RESEARCH

- **Understanding challenges and reasons for poor data recording**. Since some metrics are widely poorly recorded across many flocks (e.g. level of full pedigree recording in Merino flocks), it would be beneficial to understand 1) why it is poorly recorded, 2) explore/devise tools to increase ease of recording, and 3) design an extension campaign to target improvement of recording.
- **Cost benefit analysis and tools** to understand to what extent it is worth improving data recording, considering the costs associated, at both the individual flock and the industry levels
- **Better demonstration of the value proposition.** Flocks with higher Data Quality Scores had higher index accuracies and rates of genetic gains. The Data Quality Score also provides additional information not captured in EBV/index accuracy.
  There is potential to undertake a simulation to better understand how changing data recording reflect consequent outcomes in genetic gains.
  An important component over time will be to understand if RUGG and DQS reporting leads to change in behaviour and improved recording.
- **Better accounting of fixed effects.** Completeness and accuracy of fixed effect recording is only captured to a limited extent in the proposed framework. This requires more in-depth examination.
- **Data Quality Score -- reference vs. individual breeder.** The proposed DQS characterises the data, with the purpose of understanding the value of the data to individual breeders and their clients. Although related, an alternative perspective is valuing data contributing to the reference population, and/or other breeders.

INDUSTRY IMPLEMENTATION AND RECOMMENDATIONS

- **Public and private reporting:** The recommended roll out strategy is to initially privately report the DQS whilst road-testing, before public reporting of star ratings after a grace period. The pathway to public release (including the length of the grace period) is yet to be fully defined.
- **Continued road-testing and education:** This is particularly important if there is a reward or incentive to having a high data quality score. This requires a detailed communication strategy, which may involve media releases, and fact sheets and videos on the Sheep Genetics website.
- **Incorporation into RUGG reports:** While an interim report is available, it would be ideal to incorporate the DQS and associated features into RUGG reports. Increasing the availability and use of the RUGG reports by service providers and breeders should also be a key strategy.
- **Continuous monitoring and refinement:** The metrics require monitoring, and weights require refinement over time. This will assist in evaluation of how effective the reporting is to entice change. There is also potential to further refine the DQS reporting
- **Understanding of poor data recording:** Understanding recording challenges, and devising targeted extension messages
- **Application in beef:** The demand for an updated data quality framework for BREEDPLAN (currently delivered through DataAudit software) is unknown. In principle, frameworks for evaluating data quality should be consistent across species, primarily to facilitate extension, and potentially to simplify introduction of systems for valuing of data for reference populations.

A key contributor to success has been the collaborative nature of the project. There was consistent communication and input with MLA and Sheep Genetics from the project conception to DQS delivery. Constructive feedback from breeders has also been very beneficial for this project.

Acknowledgements

Table of contents

1. Background

The industry problem

The genetic evaluation systems for the Australian red meat industry, Sheep Genetics and BREEDPLAN, primarily rely on industry data submitted by seedstock breeders. While there are standards and guidelines for data submission, there is wide variation in data recording and submission across industry. This is a major concern for industry, as the quantity and quality of data affects the accuracy of estimated breeding values (EBVs) and genetic parameter estimates. In turn, this can lower accuracy of selection decisions and rates of genetic progress.

The accuracy (%) figure reported alongside EBVs provides producers an indication of the reliability of the estimate. The level of recording and management group structure has been shown to affect the accuracy of genetic merit estimation (Brown et al. 2001; Swan and Brown, 2007). However, EBV accuracy is calculated using the amount and structure of information utilised (i.e., quantity), and not explicitly the quality of information. The difference between data quantity and quality is highlighted in the following example: a date of birth may be supplied for each animal (maximum data quantity), but a single generic date may be used for all animals irrespective of their actual date of birth within the lambing period (poor data quality). This will affect the ability to accurately correct for age and thus the accuracy of the EBVs, but there is no way to account for this when calculating EBV accuracy. This highlights the need for additional metrics beyond EBV accuracy to characterise data quality.

Defining data quality for genetic evaluation systems

The quality of data depends on what is trying to be achieved and the intended purpose of the data. Genetic evaluation systems aim to identify the genetically superior animals to contribute as parents of the next generation. The quality of genetic evaluation data can therefore be assessed according to how well data can be used to make unbiased comparisons between animals.

To decide which animals are genetically superior, statistical analysis is conducted using data on:

- the animal's genetic makeup (through pedigree or genomic information),
- the characteristics of interest and related characteristics (i.e. performance traits and correlated traits), on individual animals and their relatives
- systematic effects that are known to affect the characteristics of interest (e.g. age, twins vs. singles)
- environmental and management conditions

In addition to requiring accurate data on the above, adequate data structure (e.g. sufficient representation of animals across systematic effects and management groups) is required to accurately estimate the genetic merit of animals (reported as estimated breeding values, EBVs).

This project developed metrics to describe data quality used in genetic evaluation systems, and an overall Data Quality Score (DQS) to rank seedstock producers on the overall quality of their data. The work focused mainly on the sheep genetic evaluation systems delivered through Sheep Genetics.

Benefits to industry

A framework to characterise genetic evaluation data will benefit individual breeders, ram and bull buyers and the industry as a whole. Feedback on the specific characteristics of their data will assist ram & bull breeders to target improvements in their recording program, to support selection decisions

and maximise genetic gains. The DQS will be a way to help identify and highlight breeders who collect high quality data. In turn, this will provide increased transparency to ram & bull buyers about the quality of data used to calculate EBVs. While EBV accuracies are available for individual animals, a DQS provides an indication of the overall quality of a flock's data and allows direct comparison across flocks. There is also potential to use these metrics to determine how data is valued and rewarded for their contribution to the reference population.

2. Project objectives

All project objectives were achieved for sheep genetic evaluation systems:

- **Objective 1. Demonstration of the value proposition** of data quality metrics in relation to prediction of genetic merit

- **Objectives 2 & 3. Development of applicable data quality metrics and refinement of the current data quality reports and metrics** within Sheep Genetics, as well as any additional metrics from BREEDPLAN

- **Objective 4. Demonstration of ways to calculate and report data quality metrics** on a per flock basis to the public, including OVIS software development

3. Methodology

This section is structured to reflect the objectives defined above.

3.1 The value proposition for current data quality metrics

The value of current data quality metrics was demonstrated by understanding their relationships to 1) genetic gains, 2) ability to predict progeny performance, and 3) genetic parameter estimates.

Current data quality metrics

All investigations were conducted using data from the Terminal (TERM), Merino (MERI) and Maternal (MATL) analyses. Current data quality metrics were computed through the routine 'stoplights' analyses conducted in routine runs, taken from the 4th January 2020 TERM analysis (441 flocks), 31st January 2020 MERI analysis (243 flocks), and 4th January 2020 MATL analysis (140 flocks).

The current data quality metrics were calculated on an individual flock basis averaged over the last 5 year period:

- **Fullped**: proportion of animals from the flock in the analyses with full pedigree (%)
- **Noped**: proportion of animals with no pedigree (%)
- **Recorded**: proportion of the year-drop that has been recorded for any of the following: weight, fat, eye muscle depth, fleece weight and fibre diameter (birth, weaning, post-weaning, hogget and adult stages) and number lambs weaned
- **Ncg**: number of contemporary groups.
- **Eff**: Average Effective Progeny. This describes how much of the information submitted is contributing to the EBV calculations for the trait, averaged across contemporary groups. This is calculated taking into account the number of progeny of each sire group compared to the total number of progeny in the group. Effective number of progeny for each sire was calculated using the formula;

$$NEF \quad = \quad \sum_j (n_j (N_j - n_j) / N_j)$$

Where NEF = Number of effective progeny, n_j = the number of progeny by the sire in group j, and N_j = the total number of progeny in group j.

The effective progeny counts are summed within each contemporary group and expressed as a proportion of the number of animals within the group. Thus the greater the value, the more effective the information is in the analysis.

- **Prop_link**: the average proportion of animals recorded that are directly linked to external flocks. This is also averaged across contemporary groups for the trait. A higher value means more linkage, and the more accurately we can predict animal performance and account for non-genetic factors (%)
- **M:F**: male to female count (ratio)
- **Avpedknown**: average completeness of pedigree known from last 3 generations (%)
- **Ngeno**: average of the proportion of animals genotyped each year (%)
- **Overall**: an index of multiple quality metrics equally weighted (includes fullped, noped, recorded, ncg, eff, prop_link; as well as metrics describing genetic gains (defined in the next section)).

3.1.1 Data quality metrics and genetic gains

This analysis aimed to demonstrate the value proposition for data quality by understanding the value of current data quality metrics in terms of observed genetic gains. The indexes examined were those that were relevant for each breed type:

- Lamb Eating Quality (LEQ) index for TERM: an index targeting both growth and eating quality for prime lamb
- Merino Production + (MP+) for MERI: for fine wool operations that target fleece weight, fibre diameter, body weight, staple length and number of lambs weaned
- Maternal Carcase Production (MCP) for MATL: for self-replacing systems with a carcase production focus

Genetic gains were quantified by index accuracy (indexacc), average index value (avindex) and index trend (indextrend):

- **Avindex**: average index value (for the relevant index of that breed type)
- **Indextrend**: the average annual change in index value (for the relevant index of that breed type)
- **IndexAcc**: the average accuracy of the index relevant for each analysis. This is influenced by the trait information submitted, with higher accuracies expected with more data submitted (%)

Genetic gains were analysed using a series of simple linear regression models:

$$y_i = \mu + \beta_1 x_i + \varepsilon$$

Where *y* was either indexacc, avindex, indextrend of flock *i*, and x_i was the flock's corresponding data quality metric (i.e. **fullped**, **noped**, **recorded**, **ncg**, **eff**, p**rop_link**, **M:F**, **avpedknown**, **ngeno**, **overall**).

Outliers were removed from data quality metrics due to potential leverage and influential points affecting results of this analysis. These were defined by 1.5 × Inter Quartile Range (IQR) below the 1st quartile (Q1) and above the 3rd quartile (Q3).

3.1.2 Data quality metrics and prediction of progeny performance

This analysis aimed to demonstrate the value proposition for data quality by understanding the relationship between data quality metrics and the ability to predict progeny performance with sire EBVs.

There were 2 steps in this analysis:

1. Regression coefficient estimates:

First, the ability of sire EBVs to predict progeny performance for each flock was quantified by regressing phenotypic progeny performance on sire EBV. Sire EBVS and progeny performance data were obtained from the TERM, MATL and MERI analyses. Sire EBVs were extracted from 2017, 2018 and 2019, along with their progeny performance records the following year (i.e. 2018, 2019 and 2020) respectively. Therefore, sire EBVs were derived independently of progeny performance being analysed. Analysis was restricted to flocks that had more than 1 sire (Table 1). The regression coefficients for each flock were averaged across the 3 years.

Table 1. Description of data structure used to regress progeny performance against sire EBVs

	Number of flocks	Average number of sires per flock	Average number of progeny per sire
TERM	214	5	435
		[range 2 – 30]	[range 4 – 2,173]
MERI	30	7	704
		[range 2-63]	[range 21 – 4,331]
MATL	19	7	583
		[range 2-38]	[14 – 2,535]

The traits examined were post weaning weight (**pwt**) and post-weaning eye muscle depth (**pemd**). Preadjusted progeny performance was analysed on an individual flock basis using the following model:

$$y_{ijkl} = \mu + CG_k + \beta_1 flock_l.sireEBV_i + \varepsilon_{ijkl}$$

where y is the preadjusted phenotype (using the standard OVIS methodology) of progeny j from sire i in flock l and contemporary group k.

Regression coefficient estimates ($\hat{\beta}_1$), which quantify how well sire EBVs predict progeny performance for each flock, were extracted from each of these models. Since progeny receive half the genetic expression from the sire, the expectation for the regression coefficient estimates was 0.5.

2. Data quality metrics as predictors of ability to predict progeny performance

The current data quality metrics were assessed as predictors of the regression coefficients. Regression coefficient estimates from the first step of analysis were cleaned to only include those between -2 and +2. For each flock, the average regression coefficient for all sires and years were taken. These were analysed in a series of simple linear regression models, with data quality metrics as predictor variables (described in section 3.1).

A series of simple linear regressions were used to understand single-variable relationships:

$$y = \mu + \beta_1 x$$

Where y is the regression coefficient estimate for each flock, μ = overall mean, x is the data quality metrics (i.e. **fullped**, **noped**, **recorded**, **ncg**, **eff**, p**rop_link**, **M:F**, **avpedknown**, **ngeno**, **overall**, **indexacc**, **avindex** and **indextrend**).

3.1.3 Data quality metrics and (within-flock) genetic parameter estimates

This analysis aimed to demonstrate the value proposition for data quality by understanding the relationship between data quality metrics and genetic parameter estimates. Genetic parameters for pemd and pwt were estimated for each flock using REML in ASReml (Gilmour et al. 2009). Flocks were only included in the analysis if there were more than 400 animals from between 2015 and 2020 (Table 2).

Table 2. Summary statistics of number of animals included in each flock in the terminal (TERM), merino (MERI) and maternal (MATL) analyses, used to estimate within-flock genetic parameters

	No of flocks	Number of animals per flock				Number of sires per flock			
		Mean	Min	Max	sd	Mean	Min	Max	sd
TERM	258	1,379	401	1,056	860	30	5	119	19
MERI	136	1,734	403	1,599	1,627	36	9	138	25
MATL	94	1,671	405	1,463	1,924	33	4	182	27

The following model was used to estimate genetic parameters for each flock:

$$y_{ijk} = \mu + CG_j + sirebreed + dambreed + animal_i + damPE_k + \varepsilon_{ijk}$$

where y is the pre-adjusted (by OVIS for the standard Sheep Genetics effects) phenotype of animal i from contemporary group (CG) j, animal is the random additive genetic variance term (quantified using pedigree) and damPE is the random permanent environment of dam k (fitted for pwt only). Phenotypes were preadjusted for sex, age, birth type, rear type and age of dam.

A series of simple linear regressions were used to understand the relationship between heritability estimates and current data quality metrics:

$$y = \mu + \beta_1 x$$

where y is the heritability estimate for each, μ = overall mean, x is the data quality metric (i.e. **fullped**, **noped**, **recorded**, **ncg**, **eff**, p**rop_link**, **M:F**, **avpedknown**, **ngeno**, **overall**, **indexacc**, **avindex** and **indextrend**).

3.2 Development and refinement of data quality metrics

The following work was conducted to refine existing data quality metrics and develop additional data quality metrics:

3.2.1 Review of the existing frameworks

Understanding past and current frameworks to characterise genetic evaluation data provides knowledge and inspiration for further refinement of data quality metrics. The following frameworks were reviewed:

- LAMPLAN Data Quality Grades
- Sheep Genetics RAMPing Up Genetic Gains (RUGG) reports
- Sheep Ireland Data Quality Index (DQI)
- Irish Cattle Breeding Federation Herd Data Quality Index
- BREEDPLAN DataAudit

3.2.2 Development of a data quality framework

This was developed after reviewing existing frameworks, discussions with MLA and Sheep Genetics extension, adoption and development officers and research scientists who use the current RUGG reports as an extension tool, and discussion with Sheep Ireland about their DQI development and industry reception.

3.2.3 Refinement of current metrics and calculation of additional metrics

Current metrics were categorised into the data quality framework components of quantity, quality and timeliness. Additional metrics that are currently calculated and reported in RUGG reports were further explored:

- **synped** (%): proportion of animals with syndicate pedigree (i.e. where multiple rams are mated over a group of ewes, resulting in multiple potential parents for the progeny).

- **ages** (%): proportion of animals recorded that are in contemporary groups with variation in age. Variation in age within contemporary groups is expected with accurate birth date recording.

- **bt** (%): proportion of animals recorded that are in contemporary groups with variation in birth type recorded.

Metrics were further refined, and new metrics were developed, including metrics from BREEDPLAN Data Audit software.

The following new and refined metrics were developed:

Quantity

- **cnt**: average flock size
- Recorded, the proportion of the year drop that has been recorded across multiple traits, was further refined into the average proportion of year drop recorded for the following trait groups:
    - **rec_weights:** bwt, wwt, pwt, ywt, hwt, awt, wcf, pcf, ycf, hcf, wemd, pemd, yemd, hemd,
    - **rec_repro:** nlw
    - **rec_wool:** pgfw, ygfw, hgfw, pfd, yfd, hfd

Quality

- **Max_freq:** % of the most common single value appearing. Missing values (designated as "-1.0") were not included in the calculation.

- Chi-squared statistics: For a reasonable sample size of data with sufficiently wide variation in values, the likelihood of any particular number appearing as a last digit is 10% (Dlugosz and Muller-Funk, 2009). This metric has been used in fraud detection by the taxation office. In the genetic evaluation context, deviation from this expectation may be due to poor recording, equipment problems or non-randomisation of recordings. Different traits are recorded and submitted in various increments (e.g. as whole number integers, or various decimal places). Therefore, the chi-squared statistic was calculated for the digits in the units place value (**chi_units**) and tenths column (**chi_tenths**). For example, for the wwt of 27.8 kg, the digit in the units place value is 7, and the digit in the tenths place value is 8 (Figure 1).
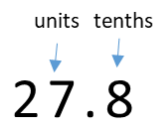
units tenths

$$2\,7\,.\,8$$

Figure 1. Example of units and tenths place values, used to calculate chi-squared statistics

A chi-squared statistic was calculated to evaluate the deviation from the expected frequencies for each digit in the units and tenths placed value:

$$\chi^2 = \sum_{i=0}^{9} \frac{(\text{expected\%} - \text{observed\%})^2}{\text{expected\%}}$$

where expected = 10%, and observed = % of records with the digit i

At a significance level of 0.05 and 9 degrees of freedom, the critical chi-squared statistic is 3.325. This means that a flock with a chi-squared statistic of less than 3.325 has recorded digits according to what is expected (i.e. 10% frequency for each digit). If a flock's chi-squared statistic is greater than 3.325, there are significant deviations from what is expected. The greater the chi-squared statistic, the greater the deviation.

All weight traits were used to calculate chi-squared statistics for digits in the units and tenths place values: birth weight (**bwt_chi_units**; **bwt_chi_tenths**), weaning weight (**wwt_chi_units**; **wwt_chi_tenths**), post-weaning weight (**pwt_chi_units**; **pwt_chi_tenths**), yearling weight (**ywt_chi_units**; **ywt_chi_tenths**), hogget weight (**hwt_chi_units**; **hwt_chi_tenths**), and adult weight (**awt_chi_units**; **awt_chi_tenths**). Sufficient records are required to adequately capture the frequency of each digit. Therefore, each flock required a minimum of 50 records for each trait across the 5 years to obtain a chi-squared statistic. Missing default values of "-1.0" were also not included in this calculation.

- Date of birth distribution:

These metrics are derived using the same concept as the chi-squared statistics. For a reasonable sample size of date of births, the likelihood of birth dates to occur on any given day of the week or month is expected to be equal. For example, the likelihood of a ewe to lamb on Monday is the same as the likelihood to lamb on Tuesday, Wednesday etc.

Note that this metric may not be applicable for artificial insemination (AI) flocks, where insemination for all ewes are on the same day and so less variation in birth dates is expected.

Nevertheless, variation in the days of the week and month for dates of birth is still expected for AI flocks.

The metrics calculated for this concept include:

- **dayinweek**: birth date deviations from expected days of the week. This is calculated as

$$\sum_{i=1}^{7} (\text{expected\%-observed\%})^2$$

where expected = 1/7 × 100 % for each day of the week, and observed= % of animals born on the *i*th day of the week.

Note that the birth date deviation from expected days in the month was also explored, with expected frequencies of 1/31 × 100 % for each day of the month. Since the correlation with dayinweek was > 0.91, results are not further discussed

- The mean linkage by trait group:
  - **link_carcase:** carcase traits of wemd, pemd, yemd, hemd
  - **link_weights:** weight trait of wwt, pwt, ywt, hwt
  - **link_repro:** reproduction traits of ynlw, nlw
  - **link_wool:** wool traits of pfd, yfd, hfd, afd

Timeliness

The new timeliness metrics developed described pedigree and trait data submission:

- **AgePed:** the average age that animals first appear in the pedigree i.e. run date – date of birth. This concept is depicted in Figure 2. These metrics were calculated by iterating through every analysis run conducted in the last 5 years.

Figure 2. Visual representation of how pedigree timeliness, quantified as age of inclusion in pedigree analysis, is calculated

- Age of trait data submission
  This is calculated as the average difference between the run date when animals have trait data and date of measurement. i.e. run date – date of measurement. This concept is depicted in Figure 3. Again, these metrics were calculated by iterating through every analysis run conducted in the last 5 years.
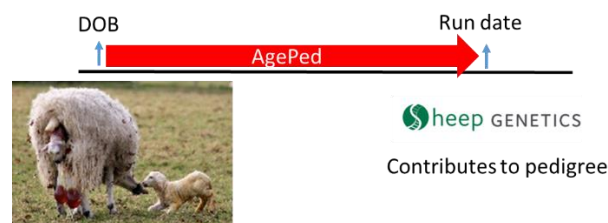
Figure 3. Visual representation of how trait timeliness data, quantified as age of trait data submission, is calculated

This was initially calculated for yearling fibre diameter and weaning weight. This was further refined as an average by trait group: carcase scan traits (**time_carcase**), reproduction traits (**time_repro**), wool traits (**time_wool**), and weight traits (**time_weights**).

The relationships between all metrics (new and existing) were explored to understand which variables capture the same underlying aspects, and therefore provide some guidance as to which metrics may be discarded to reduce complexity. Pearson's correlations were calculated to understand pairwise linear relationships between each metric. How the metrics relate to each other simultaneously was further explored in principal component analysis in Section 3.3.

Due to the inconsistent level of recording, and hence availability of chi-squared metrics, the data quality metrics of chi-squared units, chi-squared tenths metrics and maximum frequencies were not included in this analysis.

3.2.4 The Data Quality Score (DQS)

To combine the metrics explored into an overall quality of a flock's data, a Data Quality Score (DQS), was derived. To decide which metrics contributed to the DQS, and their relative importance, three different methods were explored. These were conducted on an individual breed level. Due to the different scales and variances of each data characteristic metric, each metric was scaled to unit variance and mean of zero. The resulting DQS for each method was also then scaled to between 0 and 100 for ease of interpretation.

Method 1: Principal component analysis (PCA) of all metrics (**DQS_PCA**)

PCA was conducted to select metrics used for the DQS calculation, by identifying clusters of metrics and reducing similar metrics. The principal component loadings were then explored as weightings for each metric. The first and second principal component loadings (**PC1**, **PC2**), and simultaneous use of PC1 and PC2, weighted by their relative variance explained (PC1&2) were explored. Only the use of PC1 is discussed here due to the high correlation between PC1 and PC1&2, and PC2 identifying poor data quality flocks.

Method 2: Multiple linear regression (MLR) of index accuracy (**indexacc**), average index (**avindex**) and index trend (indextrend).

There were 4 genetic gains metrics analysed, to 'train' the DQS weightings:

- Index accuracy (**DQS_indexacc**)
- Index trend (**DQS_indextrend**)
- Average index (**DQS_avindex**)
- The 1st principal component of index accuracy, index trend and average index (**DQS_GG_PC1**)

Data metrics were firstly tested for multi-collinearity (i.e. linear relationships between variables) using the variance inflation factor (James et al., 2017). A series of stepwise MLR were used for metric reduction and weight allocation (through partial regression coefficients):

$$\text{GeneticGains} = \mu + \beta_1 x_i + \varepsilon$$

where GeneticGains is **indexacc**, **avindex**, **indextrend** and **GG_PC1** of flock i and x_i the flock's corresponding data characteristic metrics. Outliers, defined by 1.5 × Inter Quartile Range below the 1st quartile and above the 3rd quartile, were removed from data metrics due to potential leverage and influential points affecting results of this analysis.

Method 3: Informed weightings (DQS_Informed)

This weighting structure was derived taking into account:

- Knowledge on the variation in the data characteristic metrics
- Relationships between metrics
- Results of PCA and MLR methods
- The breeding objectives of each breed type
- Industry knowledge and experience

Expectations on direction of weights
Weighting estimates from the above methods were expected as follows:

- Positive (i.e higher values in these metrics expect to contribute to a higher DQS score)
    - Pedigree recording and depth
    - Number of animals recorded and genotypes
    - Variation in ages and birth type
    - Effective progeny numbers
    - Linkage with other flocks
- Negative (i.e. higher values in these metrics expect to contribute to a lower DQS score)
    - Syndicate pedigree
    - Deviation from expected frequencies and distributions for traits, birth dates etc.
    - Time taken for data submission

3.3 Demonstration of data quality metric reporting

3.3.1 Data Quality Framework reporting

Data Quality Star Rating

To provide some leeway in the weights used to derive the DQS, a star rating system was created. Three methods were explored for partitioning of the DQS: a) quintiles (equal number of flocks per star rating), b) equal split in DQS range (0-20, 20-40, 40-60, 60-80,80-100) and c) normal distribution in percentiles – 10%, 20%, 40%, 20%, 10% in star ratings 1 to 5, respectively.

New software was written to automatically generate an interim data quality score report. This data quality score prototype included:

- Overall DQS
- DQS Star rating
- Where they sit compared to other flocks
- New/refined metrics – e.g. recorded and linkage by trait group, day in week, timeliness of data. These can be incorporated into existing categories (e.g. 'Pedigree Analysis') or in new categories (e.g. "Timeliness")
- Highlight data recording strengths
- Provide recommendations for where recording can be improved

Identification of strengths and recommendations

Two methods were explored to identify each flock's strengths and areas for improvement: 1) according to how they rank in percentiles for each metric, or 2) absolute deviation from the flock with the best value for each metric. Since flocks can be placed in the same percentile for multiple metrics, method 2 was the more precise and the recommended method to identify data recording strengths and recommendations for areas for improvement.

This algorithm was used across all flocks, even for high DQS flocks with a DQS 5 star rating. Therefore, if a flock that was within the top 20% of a metric, this metric was not provided as a recommendation. Note that the distributions of each metric had to be taken into consideration to determine the threshold for whether a recommendation was provided or not. For example – since **fullped** was negatively skewed with a high average of 89% for TERM flocks, the threshold was set at the median (reporting threshold of 97%) as opposed to the top 20% (reporting threshold of 99.5%).

3.3.2 Feedback and road-testing with industry

The data quality framework was presented at the AGBU Technical committee meeting on the 21st March 2021 and the Advisory Committee in April 2021. Following endorsement to proceed with the informed weightings approach, a DQS report was developed.

DQS reports were generated for 101 flocks, and in collaboration with Sheep Genetics, were road-tested at the following events:

- Service provider workshop, Melbourne, 5th May 2021 (13 flocks reports)
- Armidale Regional Forum, 19th May 2021 (8 reports)
- Launceston Regional Forum, 25th May 2021 (8 reports)
- Adelaide Regional Forum, 3rd June 2021 (16 reports)

- Merinolink ram breeder workshop, 8th - 9th June 2021 (31 reports)
- $uperborder$ Conference, 15th June(14 reports)
- Dubbo Regional Forum, 23rd June 2021 (11 reports)

Workshop participants that did not provide flock codes, or had insufficient data to generate RUGG and DQS reports, were given a generic de-identified report (MATL, TERM or MERI) for a flock selected at random.

The data quality session for each of the above workshops was organised as follows:

- Introduction to the concept of data quality in relation to the intended purpose of genetic evaluation systems and data requirements
- The data quality feedback currently provided through Sheep Genetics RUGG reports
- Introduction of the data quality score prototype
  - the development
  - proposed role within industry
  - inclusion of DQS to further enhance RUGG reports
- RUGG reports and DQS reports were provided for each flock (nominated by the participant prior to the workshop) – participants were given time to read, reflect and ask questions individually.
- Group discussion about the DQS and how it could be improved (prompts via anonymous live polling via slido)
- Filling in of anonymous feedback sheets (example provided in Appendix 1)

3.4 The value proportion for a Data Quality Score

The relationships between genetic gains and the Data Quality Score star ratings was explored to demonstrate the value proposition.

This was demonstrated by

1) plotting of index accuracies, average index and index trend for each of the DQS star rating groups
2) understanding the amount variation in genetic gains explained by the DQS

$$\text{GeneticGains} = \mu + \beta_1 DQS_i + \varepsilon_i$$

where GeneticGains is **indexacc**, **avindex** and **indextrend** for flock *i* and DQS$_i$ the flock's corresponding Data Quality Score.

4. Results

4.1 The value proposition for current data quality metrics

Figure 4 demonstrates the extent of variation in the data quality metrics currently calculated. While the distributions across the 3 analyses are overlapping, the medians of **fullped** and **avpedknown** for MERI were notably lower compared to the medians for TERM and MATL.
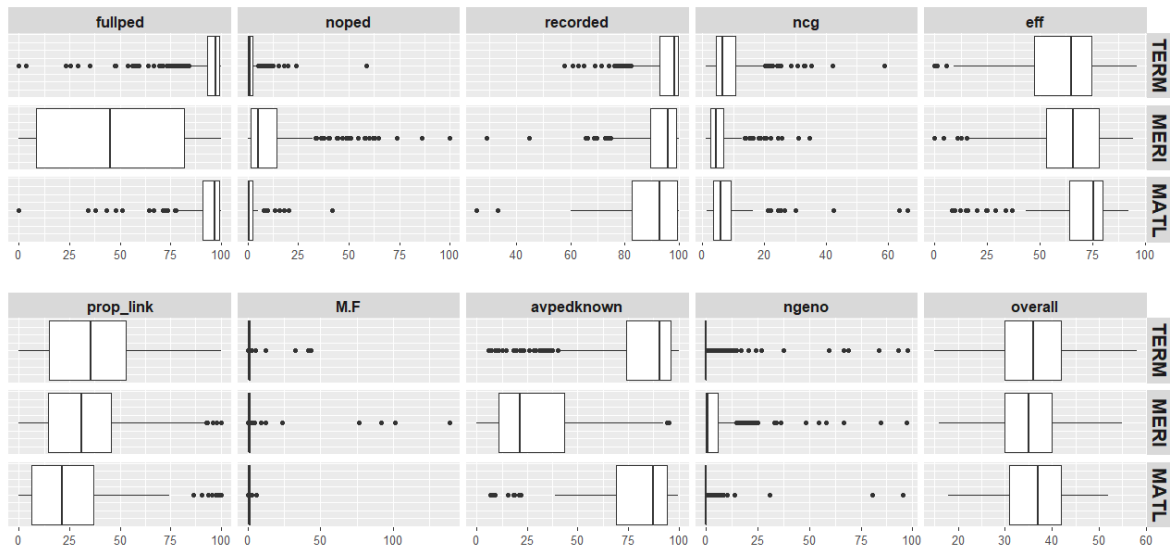
Figure 4. Distributions of current data quality metrics* for flocks in the Terminal (TERM), Merino (MERI) and Maternal (MATL) January 2020 analysis.

The genetic gains metrics were also distributed across the breeds (Figure 5).
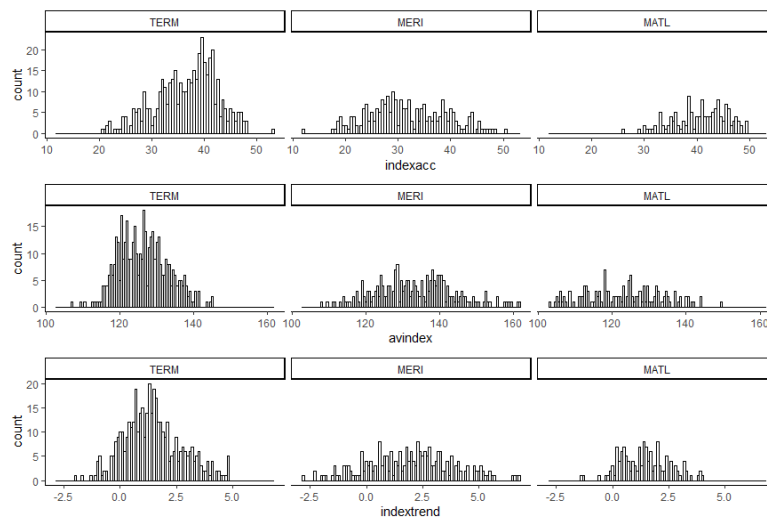
Figure 5. distributions of index accuracy (index, %), average index values (avindex, index units) and annual index trend (indextrend, index points/year) for the lamb eating quality index for Terminal (TERM) flocks, Merino Production Plus index for Merino (MERI) flocks, and Maternal Carcase Production index for Maternal (MATL) flocks in the January 2020 analysis.

Pearson's correlations were examined to firstly assess the linear relationships between the data quality metrics. There were low to moderately strong linear associations between all data quality metrics. There were strong associations between **avpedknown** and **indexacc** (0.72, 0.78 and 0.67 for for TERM, MERI and MATL flocks, respectively), **avpedknown** and **overall** (0.64, 0.69 and 0.60, respectively) and **overall** and **indexacc** (0.69, 0.8 and 0.78, respectively). However, as noted above, the overall index was calculated using the other metrics, and this part-whole relationship results in expected high Pearson's correlations. Nevertheless, these Pearson's correlations demonstrate that metrics may relate to each other and capture similar aspects of data quality. The relationships between metrics is further explored when combining all metrics into an overall data quality score (Section 4.3).

The value proposition for current data quality metrics was demonstrated by examining the relationships between current data quality metrics and 1) genetic gains, 2) regression of progeny performance on sire EBVs, and 3) genetic parameter estimates.

4.1.1 Data quality metrics and genetic gains

There were significant associations between current data quality metrics and average index merit and index trend (Table 3). Note that the large regression coefficient estimate and proportion of variation explained (R^2) by the **Overall** index is due to its part-whole relationship with index accuracy, index gains and index trend.

Most data quality metrics had a significant association with **indexacc**, **avindex** and **indextrend** for TERM flocks. Positive estimated regression coefficients suggest that **indexacc**, **avindex** and **indextrend** were significantly positively associated with:

- the number of animals with full pedigree (**fullped**)
- proportion of the drop recorded (**recorded**)
- proportion linked (**prop_link**)
- average effective progeny (**eff**)
- completeness of pedigree in last 3 generations (**avpedknown**)
- proportion of animals genotyped (**ngeno**, for **avindex** only)

Similarly, **indexacc**, **avindex** and **indextrend** were negatively associated with the proportion of animals with no pedigree.

There were fewer data quality metrics related to average index and index trend for MERI and MATL flocks. This may be a reflection on the variable breeding objectives and therefore performance recording used in these flocks compared to TERM flocks. The only predictor of genetic gains that was not significant for all breeds was **M:F**.

These results reflect Stephen et al (2018) (using data from 2017 analyses), which also reported higher genetic gains in TERM flocks with higher proportion of **fullped** known, **avpedknown** (over the last 3 years) and larger **eff**. Stephen et al (2018) also found that **indexacc** was a significant predictor of **indextrend**. It should be acknowledged that while **indextrend** and **avindex** appear to be the most appropriate metric to quantify the genetic progress of a flock, these metrics do not take into account other influences such as use of outside rams of high genetic merit. Therefore, it would be valuable to derive a data quality metric that clearly defines gains due to either a) recording within a flock, or b) recording outside the flock. However, this may not be possible due to the reliance of data collected on linked flocks to estimate breeding values.

Table 3. The relationship index accuracy, average index and index trend[*] and data quality metrics[#] using linear regression. Significant predictors are provided with a regression coefficient ($\hat{\beta}_1$), standard errors in parentheses and coefficient of determination (R^2). Metric abbreviations available in Section 3 of this report. Non-significant (P>0.05) regression coefficients are not shown.

		Full_ped	Noped	recorded	Ncg	Eff	Prop_link	M.F	Avped known	ngeno	Overall
INDEX ACCURACY											
TERM	$\hat{\beta}_1$ (SE)	0.58 (0.08)	-1.01 (0.23)	0.34 (0.07)	0.47 (0.06)	0.09 (0.01)	0.12 (0.01)		0.27 (0.01)	9.31 (1.96)	0.52 (0.02)
	R^2	0.12	0.05	0.07	0.11	0.10	0.28		0.51	0.06	0.56
MERI	$\hat{\beta}_1$ (SE)	0.14 (0.01)	-0.5 (0.05)		0.47 (0.17)	0.21 (0.02)	0.14 (0.02)		0.25 (0.01)	0.49 (0.13)	0.82 (0.04)
	R^2	0.47	0.29		0.03	0.26	0.168		0.60	0.06	0.64
MATL	$\hat{\beta}_1$ (SE)	0.29 (0.09)	-1.27 (0.32)	0.20 (0.04)	0.49 (0.11)	0.14 (0.03)	0.09 (0.02)		0.16 (0.02)		0.51 (0.04)
	R^2	0.08	0.11	0.165	0.14	0.12	0.10		0.27		0.56
AVERAGE INDEX											
TERM	$\hat{\beta}_1$ (SE)	0.31 (0.10)	-0.63 (0.28)	0.23 (0.08)	0.42 (0.08)	0.05 (0.02)	0.09 (0.01)		0.20 (0.02)	6.05 (2.23)	0.45 (0.03)
	R^2	0.02	0.01	0.02	0.07	0.02	0.12		0.19	0.02	0.32
MERI	$\hat{\beta}_1$ (SE)		-0.28 (0.09)			0.13 (0.04)	0.18 (0.03)			0.55 (0.19)	0.60 (0.09)
	R^2		0.04			0.04	0.08			0.04	0.17
MATL	$\hat{\beta}_1$ (SE)				0.87 (0.23)		0.15 (0.05)				0.69 (0.10)
	R^2				0.10		0.07				0.25
INDEX TREND											
TERM	$\hat{\beta}_1$ (SE)	0.05 (0.02)	-0.17 (0.05)	0.05 (0.01)	0.06 (0.01)	0.01 (0.003)	0.01 (0.002)		0.03 (0.004)		0.09 (0.006)
	R^2	0.02	0.03	0.03	0.04	0.03	0.06		0.10		0.34
MERI	$\hat{\beta}_1$ (SE)	0.02 (0.003)	-0.05 (0.02)				0.01 (0.006)		0.03 (0.005)	0.07 (0.04)	0.12 (0.02)
	R^2	0.11	0.04				0.03		0.11	0.02	0.21
MATL	$\hat{\beta}_1$ (SE)			0.02 (0.009)							0.06 (0.01)
	R^2			0.05							0.18

[*]indexes used were lamb eating quality index for TERM flocks, Merino Production Plus index for MERI flocks, and Maternal Carcase Production index for MATL flocks in the January 2020 analysis.

[#]**fullped**: % with full pedigree; **noped**: % with no pedigree; **recorded**: % recorded for list of traits; **ncg**: number of contemporary groups; **eff**: average effective progeny; **prop_link**: % linked to external flocks; **M.F**: male to female ratio; **avpedknown**: % with 3 generations of pedigree ; **ngeno**: % genotyped; **overall**: an index of multiple quality metrics and genetic gains metrics. Full definition provided in section 3.3 of this report.

Interestingly, having more contemporary groups (**ncg**) also predicted an increase in **indexacc** for all breeds. This may potentially be a reflection of these breeders recording contemporary groups more accurately. However, the current advice provided to breeders is that having more contemporary groups could result in smaller group sizes therefore less ability to accurately compare animals. This assumption of smaller group sizes results in less effective data can be removed if a measure of contemporary group sizes, and sire representation across contemporary groups, is captured as a data quality metric. In addition to this, the consistency of contemporary group numbers may be a metric of interest, under the assumption that mob management does not change considerably over time. Therefore, this current **ncg** metric can be further refined by taking into account size of contemporary groups and change in contemporary groups over time.

Currently the **recorded** metric reflects the proportion of animal in the drop with an observation for at least one trait in the trait list specified. This is currently the same for all breed types but would be better if customised to reflect traits important to the breeding objectives of the breed.

While there was only a small proportion of the variation explained by the data quality metrics, this analysis demonstrates that 1) the current data quality metrics demonstrate the large amount of variation in the quality of data submitted, and 2) data quality metrics play an important role in determining potential rates of genetic gain. Therefore, there is value in providing this feedback to breeders so they can identify areas for improvement to maximise genetic gain.

There are also opportunities to further refine these current data quality metrics (for example, contemporary group numbers and structure as discussed above). These current data quality metrics also do not capture the timeliness of data submission and the variability (or stability) of data captured over time. In addition to this, these metrics can more appropriately reflect the specific traits of importance to the different breed types and their breeding objectives. It is recommended that the current data quality metrics be further refined, and a new overall metric developed (independent of average index and index trend) and refined weightings.

4.1.2   Data quality metrics and prediction of progeny performance

Step 1. Progeny performance was pre-adjusted using the standard OVIS systematic effect adjustments and then regressed on sire EBVs. These raw regression coefficients exhibit a lot of noise; this may be a reflection of not taking into account use of home-bred versus outside rams in this analysis, as well as inability to account for selection or potential preferential mating. Nevertheless, these extreme values were not frequently observed, and as extreme observations are known to have a significant influence on the results due to leverage, and thus these outliers were removed from the next step of analysis. The regression coefficients used were restricted to between -2 and +2 as shown below (Table 4 and Figure 6). This only removed 2 flocks from TERM.

Table 4. Summary statistics for cleaned regression coefficient estimates from progeny performance regressed on sire EBVs for terminal (TERM), merino (MERI) and maternal (MATL) analyses

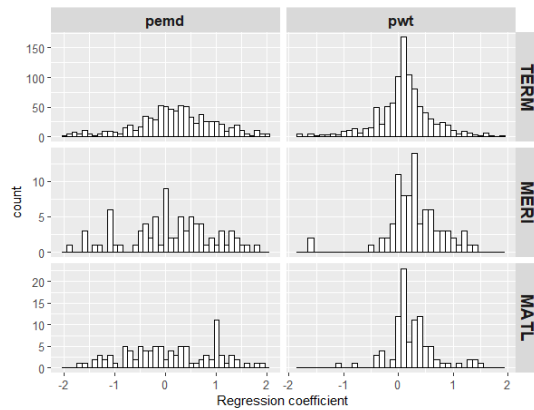|	Mean	Min	Max	SD
		pemd		
TERM	0.19	-2.00	2.00	0.77
MERI	0.10	-1.9	1.84	0.84
MATL	0.07	-1.75	1.88	0.85
		pwt		
TERM	0.11	-1.79	1.85	0.50
MERI	0.32	-1.59	1.37	0.49
MATL	0.07	-1.75	1.88	0.85

Figure 6. Distributions of cleaned regression coefficient estimates from progeny performance regressed on sire EBVs for terminal (TERM), merino (MERI) and maternal (MATL) analyses

Step 2. Simple linear regression was used to explore the relationship between data quality metrics and ability to predict adjusted progeny performance (Table 5). It should be noted that the significant relationships are largely due to high leverage. For example, the relationship between **M:F** and regression coefficients for **pemd** in TERM may be due to the one influential point of low **M:F** value.

Table 5. The relationship between ability to predict progeny performance using sire EBVS and data quality metrics using simple linear regression. Significant predictors are provided with regression coefficient ($\hat{\beta}_1$, SE), standard errors in parentheses and coefficient of determination (R^2). Metric abbreviations available in Section 3 of this report.

		Full_ped	Noped	recorded	Ncg	Eff	Prop_link	M:F	Avpedknown	ngeno	Overall	IndexAcc	AvIndex	IndexTrend
							pemd							
TERM	$\hat{\beta}_1$ (SE)							1.55 (0.74)						
	R^2							0.024						
MERI		No sig. predictors												
MATL		No sig. predictors												
							pwt							
TERM	$\hat{\beta}_1$ (SE)					0.004 (0.002)	-0.004 (0.001)							
	R^2					0.02	0.03							
MERI	$\hat{\beta}_1$ (SE)												0.02 (0.006)	
	R^2												0.23	
MATL	$\hat{\beta}_1$ (SE)	-0.04 (0.02)												
	R^2	0.28												

This analysis shows that there is little evidence that there is a relationship between data quality metrics and the ability to predict progeny performance with sire EBVs. The ability to understand the relationships between current data quality metrics and ability to predict progeny performance with sire EBVs may be limited by the number of flocks used in the analysis, the variable breeding objective for each individual flock and the amount of recording to reflect the breeding objectives. Future investigations can examine alternative traits, which may be more relevant for the different breeds. As discussed in Section 5.1, there are opportunities to develop new data quality metrics, or refine the current data quality metrics to more appropriately reflect the recording strategies of each flock. These new metrics can then be explored to understand its value to prediction of genetic merit.

There was evidence of non-linear relationships between data quality metrics (results available on request). This was explored to understand if there was an optimum data quality metric value that allows prediction of progeny performance as expected. For example, for MATL, an overall measure of 40 to provide an expected regression coefficient of 0.5. Again, since this analysis did not take into account use of home-bred versus outside rams in this analysis, selection or potential preferential mating, these results should be interpreted with caution.

4.1.3   Data quality metrics and (within flock) genetic parameter estimates

Genetic parameters were estimated on an individual flock level, using post-weaning eye muscle depth and post-weaning weight as examples. Figures 7 and 8 demonstrate that there was variation in all genetic parameters, and that there were overlap across the 3 breed types. Summary statistics for these genetic parameter estimates are provided in Table 6.
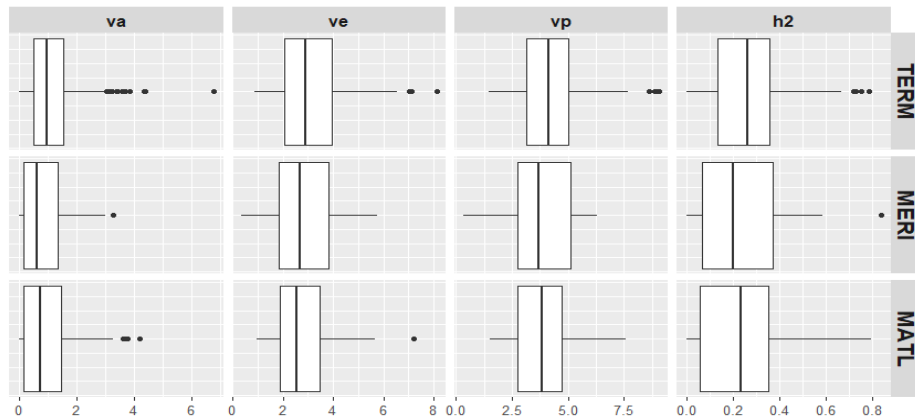

Figure 7. Within-flock genetic parameter estimates for post-weaning eye muscle depth, for terminal (TERM), merino (MERI) and maternal (MATL) breeds.

Abbreviations: va: additive genetic varianc; ve: residual variance; vp: phenotypic variance; h2: heritability estimate
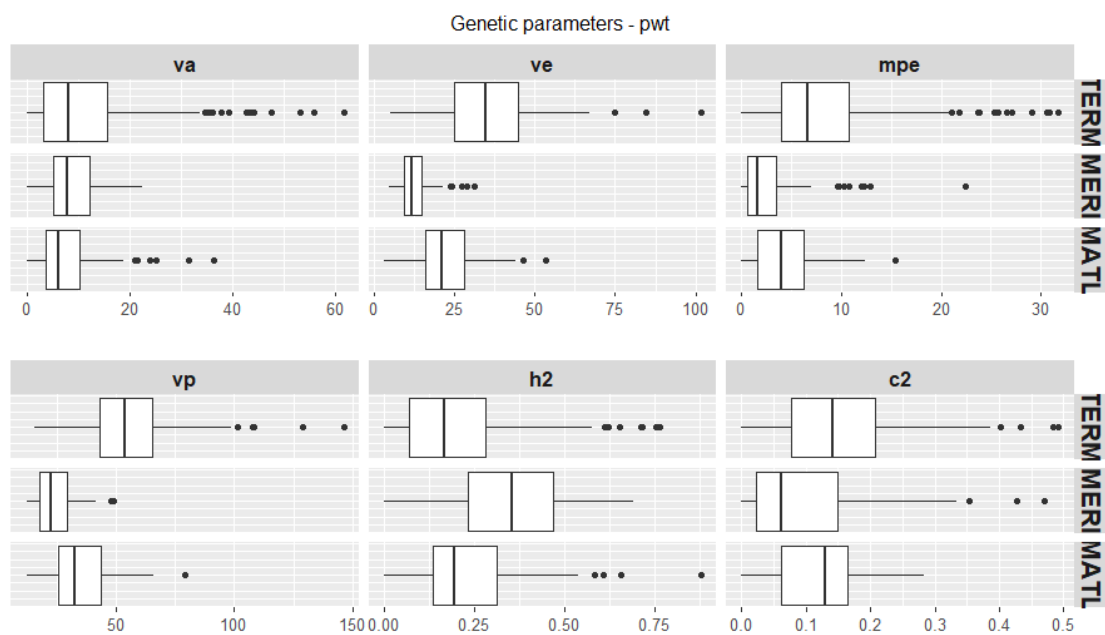

Figure 8. Within-flock genetic parameter estimates for post weaning weight, for terminal (TERM), merino (MERI) and maternal (MATL) breeds.

Abbreviations: va: additive genetic variance; ve: residual variance; mpe: maternal permanent environment effect; vp: phenotypic variance; h2: heritability estimate; c2: permanent environmental effect of the dam

Table 6. Summary statistics of genetic parameter estimates for post weaning eye muscle depth (pemd) and post weaning weight (pwt), calculated on a within flock basis for 336 terminal flocks (TERM), 156 merino flocks (MERI) and 111 maternal flocks (MATL).

	pemd																			
	va				ve				vp				h²				c²			
	Mean	Min	Max	sd	Mean	Min	Max	sd	Mean	Min	Max	sd	Mean	Min	Max	sd	Mean	Min	Max	sd
TERM	1.14	0.00	6.80	0.96	3.09	0.88	8.11	1.28	4.23	1.46	9.05	1.47	0.26	0.00	0.79	0.17				
MERI	0.89	0.00	3.28	0.87	2.86	0.36	5.74	1.35	3.74	0.36	6.28	1.41	0.23	0.00	0.84	0.20	-			
MATL	0.99	0.00	4.18	1.00	2.79	0.97	7.23	1.16	3.77	1.50	7.59	1.31	0.24	0.00	0.80	0.20				
	pwt																			
TERM	11.27	0.00	61.62	11.33	35.95	5.11	101.56	14.35	55.35	15.55	146.23	19.15	0.20	0.00	0.76	0.17	0.15	0.00	0.49	0.10
MERI	8.53	0.00	22.40	4.75	12.63	4.64	31.34	4.91	23.80	12.49	48.95	7.51	0.35	0.00	0.69	0.16	0.10	0.00	0.47	0.10
MATL	8.00	0.00	36.35	6.82	22.68	3.31	53.57	9.54	34.86	12.29	79.17	13.35	0.23	0.00	0.88	0.16	0.12	0.00	0.28	0.07

There was considerable variation in genetic parameter estimates for each flock. For **pemd**, additive genetic variance, residual variance, phenotypic variance and heritability estimates align across the different breeds. Variability of estimates were also comparable. Compared to estimates in literature, the average within-flock heritability estimate for MERI reflects the across-flock estimate of 0.24 (0.04), reported for all flocks in the MERI analysis (Huisman et al, 2008). The average within-flock heritability estimate for TERM flocks are marginally lower than the estimate of 0.32 (0.01) for TERM flocks reported by Brown et al (2015).

For **pwt**, heritability estimates were higher for MERI due to the lower phenotypic variances, although the permanent environmental effect of the dam was consistent across all breeds. These results reflect the estimates reported by Huisman et al. (2009), where across-flock heritability estimate of 0.76 (0.05) was reported for flocks in the MERI analysis. For c², Huisman et al. (2008) reported an estimate of 0.23 (0.06) when all available animals (with and without full pedigree) were included in this analysis, compared to a c² estimate of 0.04 (0.05) if only animals with known sire and dams were included in analysis.

TERM flocks had more variation in all variance estimates compared to other breed types. However, the proportion of phenotypic variation attributed to genetic variation (i.e. h²) for TERM was similar to MATL. The lower **pwt** heritability estimates for TERM reflect the 0.17 (0.01) estimate reported by Brown et al (2015).

While there are some noticeable differences across breeds, the overlapping distributions indicate that there are no significant differences in genetic parameter estimates for each trait across the different breeds.

Simple linear regression was used to explore the relationship between data quality metrics and genetic parameter estimates (Table 7). While some current data quality metrics were statistically significant predictors of within-flock heritability estimates, these effects were extremely small. There was also no metric that was consistently of importance across breed types.

Table 7. The relationship within-flock heritability estimates and data quality metrics using simple linear regression. Regression coefficient estimates ($\hat{\beta}_1$) and coefficient of determination (R^2) provided for significant predictor variables

		Full_ped	Noped	recorded	Ncg	Eff	Prop_link	M.F	Avpedknown	ngeno	Overall	IndexAcc	AvIndex	IndexTrend
							pemd							
TERM	$\hat{\beta}_1$ (SE)				-0.006 (0.002)									
	R^2				0.03									
MERI	$\hat{\beta}_1$						No significant predictors							
	R^2													
MATL	$\hat{\beta}_1$ (SE)			-0.003 (0.001)			-0.002 (0.0007)						0.003 (0.001)	
	R^2			0.048			0.054						0.070	
							pwt							
TERM	$\hat{\beta}_1$ (SE)				-0.005 (0.002)		-0.001 (0.0005)			0.009 (0.004)			-0.003 (0.001)	
	R^2				0.023		0.025			0.055			0.048	
MERI	$\hat{\beta}_1$					0.002 (0.001)								
	R^2					0.040								
MATL	$\hat{\beta}_1$						No significant predictors							
	R^2													

Non-linear relationships were also explored (table and graphical summaries available on request). Again, while there were few metrics had significant relationships with within-flock heritability estimates, these effects were extremely marginal and inconsistent across breed types.

4.2 Development and refinement of current data quality metrics

4.2.1 Data quality framework requirements

A review was conducted on the existing frameworks available to report the data quality for genetic evaluation systems (available on request), in particular the review of the Sheep Genetics Data Quality Grades by Martin (2014). Following discussions with MLA and Sheep Genetics extension, adoption and development officers and research scientists who use the current RUGG reports as an extension tool, an updated framework to report data quality metrics should take into account the following aspects:

- Clear objectives about why the data is being characterised, and how the framework will be used
- Transparency about the scientific rigour behind how the metrics are calculated and its value to genetic gains
- Relevance to the breeding objectives of each breed type
- A dynamic, flexible framework with metrics that are easily interpreted by industry
- Interpretable metrics, realistic actions, goals and benchmarks, as well as tools to achieve them

4.2.2 Proposed data quality framework

The objectives of this proposed data quality framework are to:

- Understand the data submitted, and characterise its usefulness to genetic evaluation systems. The benchmarking of the quality is important with the increase in phenotypes and genomic data.
- Identify how breeders can improve data submission to improve potential genetic gain
- Identify and highlight the breeders who make the effort to collect good data. This may be useful to decide whether a breeder is selected to contribute to research programs, or contribute to the reference population
- Provide increased transparency to ram buyers about the data used to calculate EBVs

Besides deciding breeder contribution to research programs or reference populations, this framework and the metrics calculated can be used internally for checking data pre-analysis ('internal policing'). These metrics may also be helpful in determining how data contributing to the reference population is valued and rewarded.

The proposed Data Quality framework is depicted in Figure 9. There are 3 main components— quality, quantity and timeliness (defined below). Each circle represents one component that characterises data, and the current stoplight metrics calculated that best fit each component. The size of the circles reflect the levels of importance to estimating genetic merit (further discussed below). The overlapping of circles acknowledge that there are difficulties in developing metrics that clearly separate categories.

Figure 9. A proposed Data Quality framework to characterise data used in sheep genetic evaluation systems

The components can be described as follows:

- **Quantity**

  How much data is submitted, how much data is being used, and completeness of data submitted.

- **Quality**

  How good is the data for its intended use? This component attempts to capture how accurate the data is, as well as the data structure required for genetic evaluation. That is

  - Accuracy of the data submitted (i.e. pedigree, genotypes, phenotypes and fixed effects (including contemporary group))
  - Data structure for unbiased evaluation of genetic merit, including
    - sire representation and linkage across contemporary groups, within flock, and across flocks
    - unbiased selection of recording (partial recording and submission)

- **Timeliness**

  When or how soon after data collection the data is submitted.

  The time between data collection and submission may provide an indication of the level of engagement by flocks and breeders.

4.2.3 Refined and additional metrics

The following metrics were calculated to measure the quantity, quality and timeliness of each flock's data. This results section demonstrates variation across flocks.

Quantity metrics

Figure 10 demonstrates the variation in these new and refined data quantity metrics.
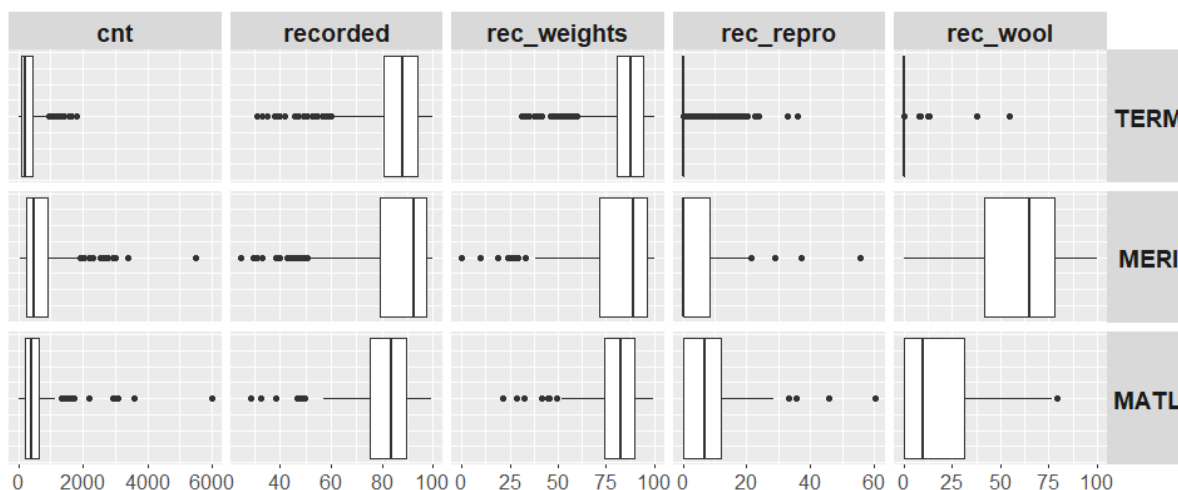

Figure 10. New and refined data quantity metrics for flocks in the terminal (TERM), merino (MERI) and maternal (MATL) analyses.

The average flock size (**cnt**) was largest in MERI (average of ~700 animals) compared to MATL (~ 500) and TERM (~300). MERI and MATL flocks were also the most variable in size (SD of 640 and 726, respectively) compared to TERM flocks (SD of 270).

The previously calculated **recorded** metric (which included all traits of interest) showed similar averages (~93%) and variation (SD of ~9%) averages across the three breeds. When the level of recording was classified by trait group, there were more distinct differences across breeds. While there were overlapping distributions across breed, the proportion of animals recorded for each trait group aligned with the general breeding objectives of each breed type:

- All breed types had a high proportion of animals recorded for weight traits (**rec_weights**). MERI flocks exhibiting the lowest proportion of weight trait recordings and also the most variability across flocks within the breed type (mean ± SD of 86.8 ± 19.1% for MERI, 94.9 ± 8.2% for TERM, and 94.2 ± 9.6% for MATL). The high proportion of weight trait recordings across all breeds may be due to weights being relatively easy to record compared to other traits.
- MATL had a greater average proportion of animals with reproduction traits recorded (**rec_repro**) and less variation compared to other breed types (mean of 8.2%, compared to 3.6% for MERI and 1.7% for TERM).
- MERI had a greater proportion of animals with wool traits recorded (**rec_wool**) (mean of 69.3%, compared to 0.36% for TERM and 15.3% for MATL), which is as expected. There was also more variation in **rec_wool** for TERM and MATL flocks compared to MERI flocks.

These results show that a quantity metric that describes the level of recording for the different breeding objectives/trait groups is more appropriate than a metric that covers a generic list of traits. Therefore, different data quantity metrics may be required for each breed type.

Quality metrics

The variation in these refined and additional quality metrics are demonstrated in:

- Figure 11:
  - syndicate pedigree (**synped**)
  - variation in age (**ages**)
  - variation in birth type (**bt**)
  - average linkage by trait group (**linkage_CA**, **linkage_WT**, **linkage_RP**, **linkage_FW**)
  - birth date days of the week (**dayinweek**)
- Figure 12:
  - chi-squared statistics – for digits in the units and tenth place value, for all weight traits recorded by each breed.
- Figure 13:
  - maximum frequency (**maxfreq**) of a single value, for weight traits recorded by each breed.

All of these additional data quality metrics demonstrated considerable variation, and there were also distinct differences between breeds.
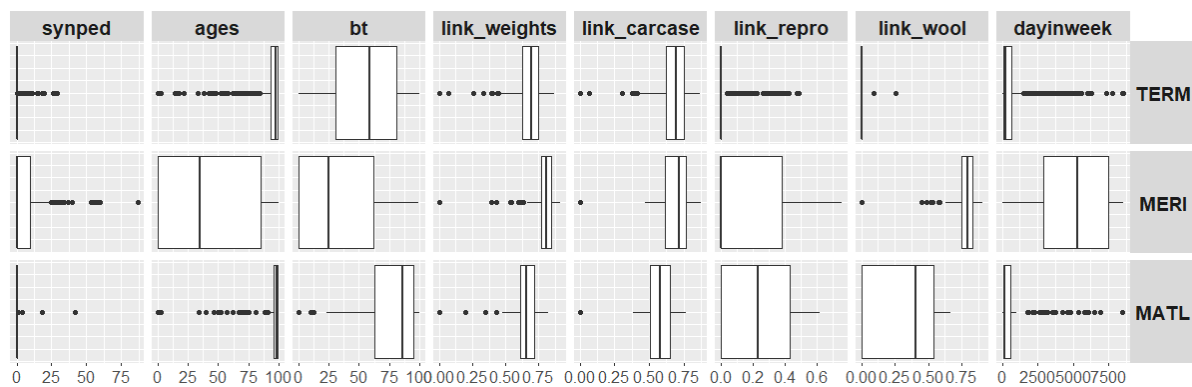

Figure 11. New and refined data quality metrics for flocks in the terminal (TERM), merino (MERI) and maternal (MATL) analyses.

Figure 11 shows that MERI flocks did not have as much variation in age (**ages**) and birth type (**bt**) recordings compared to the other breed types. There was also more variability within MERI flocks. On average, MERI flocks had 39.4±27.4% (mean±SD) of animals in contemporary groups (**CGs**) that had variation in **ages**, and 31.0±32.1% of animals in CGs with variation in birth type recordings. In comparison, MATL flocks average for **ages** was 79.4±15.7% and 73.7 ± 23.9% for bt, and TERM flocks (**age**s: 75.5±17.7%, **bt**: 53.2 ±29.0%). This suggests that date of birth and birth type is not generally as well recorded in MERI flocks.

There was also more variation in **synped** in MERI compared to MATL and TERM (SD 12.3% vs. 4.6% and 2.8%, respectively). The greater amount of syndicate pedigrees used in MERI also reflects the amount of pedigree recorded (**fullped**) above. Therefore, the quantity and quality of pedigree recording in MERI in general is below that of TERM and MATL flocks based on these metrics.

The amount of linkage to other flocks varied depending on trait group. The average linkage by carcase traits (**linkage_carcase**) and weight traits (**linkage_weights**) were similar across breeds. Meanwhile, the amount of linkage for reproductive traits (**linkage_repro**) was on average higher for MATL flocks, with MATL and MERI flocks exhibiting a large amount of variation (SD 0.22 and 0.20, respectively) compared to TERM flocks (SD 0.11%). Linkage for fleece weight traits (**linkage_wool**) was highest for

MERI flocks. These results were as expected due to the breeding objectives for each breed type. Similar to the quantity metrics by trait groups, these results show that a quality metric that describes the data structure and ability to use other flock information to derive breeding values should take into account the different breeding objective/trait groups.

The days of the week metrics provide an indication of how much birth dates deviated from expectation. On average, MERI flocks exhibited a greater deviation from what was expected for birth dates recorded in the week, which reflects the **ages** metric. Since MERI flocks had date of birth records less evenly distributed across the weekly time periods, this indicates lower accuracy of actual date of birth recording in these flocks.

The chi-squared statistics calculated provide an indication of how much digit frequencies deviate from what is expected (Figure 12). Chi-squared statistics were calculated for the digits in the units place value, as well as the tenths place value, to account for the potential rounding of observations.
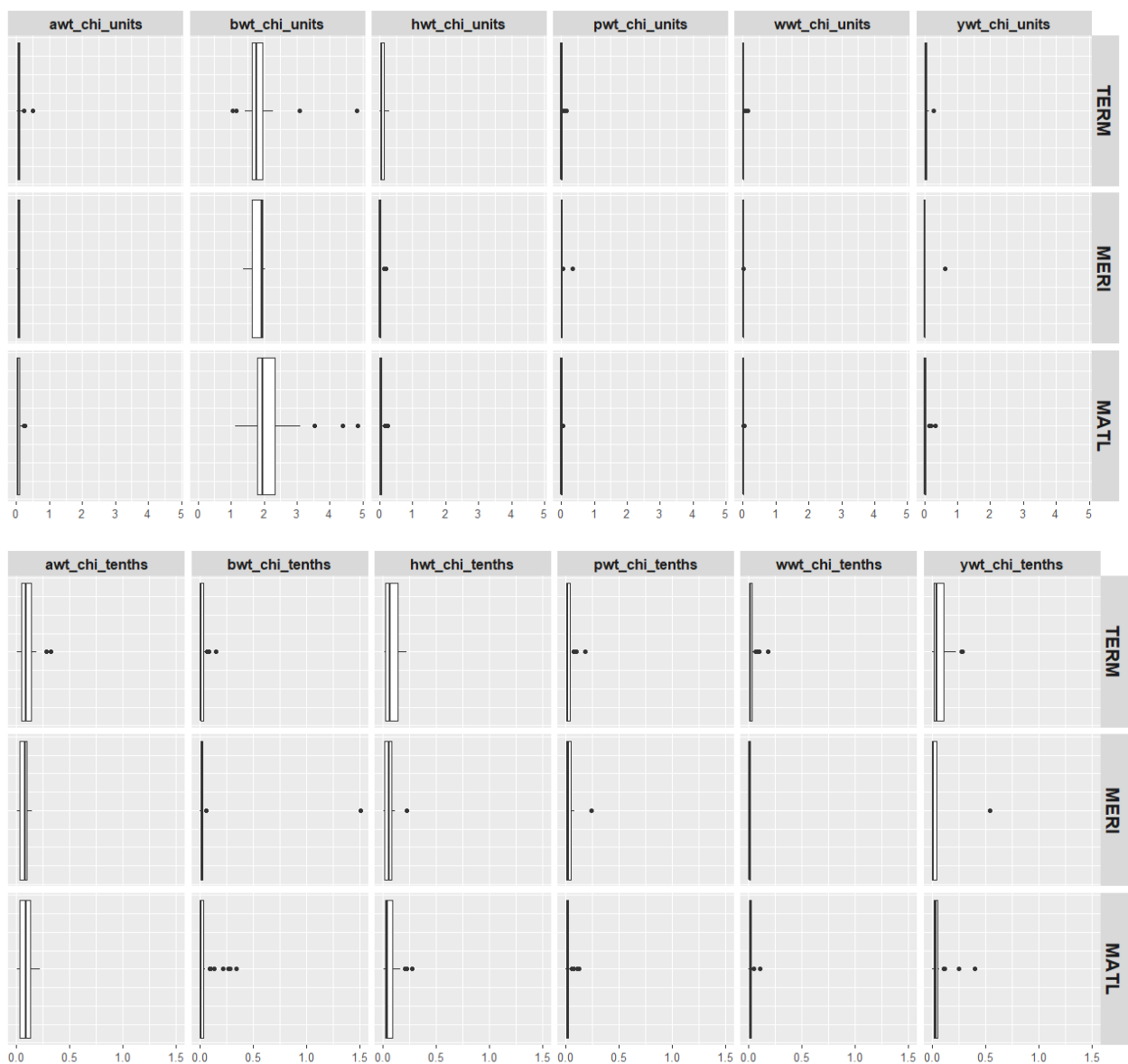

Figure 12. New and refined data quality metrics, expressed as a chi-squared statistic for digits recorded in the units (top plot) and tenths (bottom plot) place values, for flocks in the terminal (TERM), merino (MERI) and maternal (MATL) analyses

The trait that had the most variation in deviation from expected digit frequencies in the units place value was **bwt** (**bwt_chi_units**), particularly for MATL flocks. This is due to **bwt** mostly being single digit (with an average of ~4.5 kg). Therefore, the expected frequency of 10% for each digit appearing for bwt was not appropriate.

The majority of chi-squared values for all traits were less than the chi-squared critical value of 3.325, which suggests that the frequency of digits recorded for these weight traits were as expected. Therefore, this metrics appears to have limited value as feedback to breeders, but can still be used as a way to highlight any suspicious data with repeated digits.

The maximum frequency of any one observation is another metric that may be used to detect any suspicious repeated recording (Figure 13).
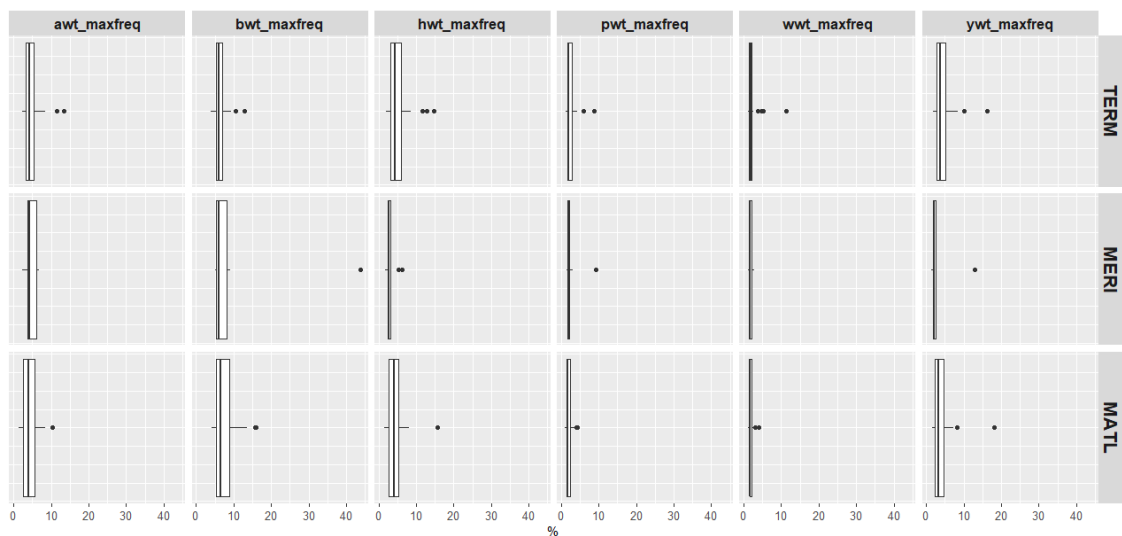

Figure 13. New data quality metrics,, expressed as the maximum frequency (%) of a single value appearing, for flocks in the terminal (TERM), merino (MERI) and maternal (MATL) analyses.

On average, in MERI flocks, 10% of records were the same for **bwt** and 12% of records was the same for ywt. MATL flocks had an average of 8% of the same records for **bwt** and 10% of the records were the same for ywt. All but 1 flock had a maximum frequency of under DataAudit's 25% 'threshold' for reporting maximum frequencies. There was 1 MERI flock that had 43% of **bwt** records of 4.4kg.

Timeliness metrics

All timeliness metrics exhibited variation between flocks (Figure 14). On average, the age at which animals first appeared in the pedigree was highest in MERI flocks (average of 406 days, compared to 268 days in TERM flocks and 255 days in MATL flocks).
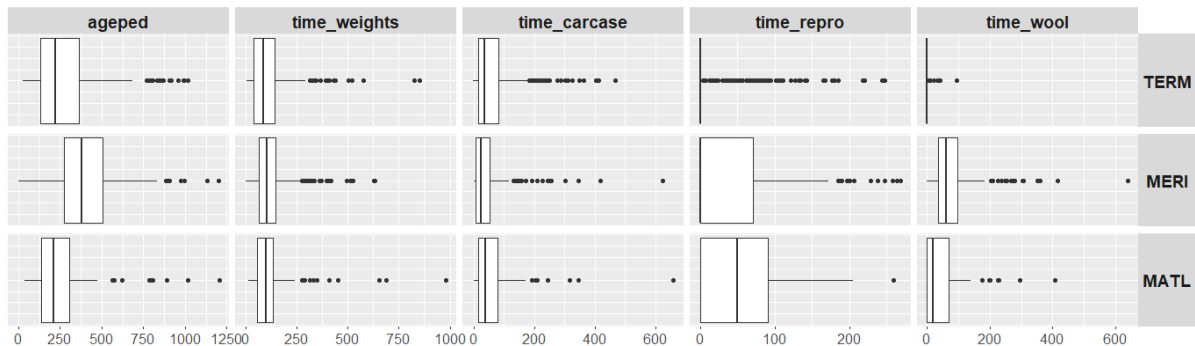

Figure 14. New and refined data timeliness metrics, describing timeliness of pedigree and trait data submission, for flocks in the terminal (TERM), merino (MERI) and maternal (MATL) analyses.

There was large variation in the amount of time taken to submit trait data after measuring. All breed types took the same amount of time to submit weight data (**time_weights**) (average of 111 days to 126 days across breeds) and also carcase data (range 43 days to 64 days). Reproduction data (**time_repro**) was submitted earlier for TERM flocks (mean 14 days) compared to MERI (38 days) and MATL flocks (54 days). On average MATL flocks submitted wool trait data (**time_wool**) earlier (44 days) compared to MERI flocks (82 days). The time taken to submit wool trait data was also more variable for MERI flocks compared to MATL flocks (SD 78 days compared to 63 days, respectively).

Relationships between metrics

Pearson's correlations were calculated to understand pairwise linear relationships all metrics. Only the Pearson's correlations (r) greater than 0.5 are outlined in this report (other relationships available on request).

Between Quantity metrics

- The generic **recorded** metric was highly correlated with **rec_weights** for TERM (r = 1.00) and MATL (r = 0.99). The strength of this linear relationship was less for MERI (r = 0.61). **Recorded** was also highly correlated to **rec_wool** in MERI (r = 0.61)
- The degree of full pedigree (**fullped**) was highly correlated with depth of pedigree (**avpedknown**) (0.73 in TERM to 0.80 in MERI).
- The proportion of animals with no pedigree (**noped**) was also moderate to highly negatively correlated with **fullped** and **avpedknown** (ranging from -0.52 to -0.75).
- The number of contemporary groups (**ncg**) reflect overall flock size (**cnt**) for all breeds, with r ranging from 0.58 for MERI to 0.80 for MATL.

3Between Quality metrics

- Linkage of weight traits (**linkage_weights**) were highly correlated to linkage for carcase traits (**linkage_carcase**) for MATL (r = 0.80) and TERM (r = 0.80)
- The amount of linkage of wool traits (**link_wool**) was also moderately correlated to **link_weights** in MERI (r = 0.53)
- In all breed types, **ages** was negatively correlated with **daysinweek** (ranging from -0.67 in TERM to -0.82 in MERI). This is as expected, as less variation in **ages** also reduces the likelihood of variation in days on the week/month when birth dates are recorded.
- **Ages** was moderately correlated with **bt** in MERI (r = 0.51) and MATL flocks only (r = 0.53)

Between Timeliness metrics

- Early submission of pedigree data (**ageped**) was moderate to highly correlated with the time taken to submit weight trait data (**time_weights**), for all 3 breeds (range r = 0.53 – 0.79). This is as expected as weights would generally be submitted at the same time as pedigree.
- The time taken to submit weight (**time_weights**) and wool traits (**time_wool**) was also moderately high for MERI (r = 0.64), and also between weight (**time_weights**) and carcase traits (**time_carcase**) in TERM (r = 0.61)

Between all metrics across components

There were also moderately strong relationships across components.

- **Fullped** had moderately strong relationships between **ages** (r = 0.71 for MATL, r =0.56 for MERI) and **bt** (r = 0.52 for MATL, r = 0.62 for MERI)
- The level of recording (**rec_repro**), time taken to submit (**time_repro**) and linkage accuracy of reproduction (**link_repro**) were all highly correlated for TERM (range 0.69-0.64) and MERI (0.68-0.72).

These pairwise analyses of metrics shows that 1) there are opportunities to minimise the number of metrics used in this proposed framework; and 2) the metrics quantifying quantity, quality and timeliness may not be completely independent of each other. Multiple metrics need to be considered simultaneously to better understand how the metrics relate to each other.

4.2.4 The Data Quality Score (DQS)

Current data quantity and quality metrics were refined, and additional metrics were developed (characterising data quantity, quality and timeliness). These metrics were related to each other and potentially could be capturing similar data characteristics. Therefore, the approach of this current milestone was to develop an overall DQS score that simultaneously encompassed all data quantity, quality and timeliness metrics (i.e. across all components, as opposed to within each component, as depicted in Figure 15). Combining all metrics across all components into a single value score is also easier to interpret compared to 3 separate data quantity, quality and timeliness scores, which would then need to be further combined into an overall score.

Figure 15. Schematic of development of the overall Data Quality Score: the metrics used and weightings for each metric

Three options were explored to decide which data characteristic metrics contributed to the overall Data Quality Score, and their level of importance (weightings).

Method 1: Principal Component Analysis (DQS_PCA)

There were 347 Terminal, 44 Maternal and 214 Merino flocks that were used to derive PCA loadings. The first principal component (PC1) explained between 25% to 30% of the total variation (Table 8).

Table 8. Principal component (PC) loadings (% of variance explained) explored as metric weights for calculation of the overall Data Quality Score, trained on 347 Terminal, 44 Maternal and 214 Merino flocks

TERMINAL (*n* = 347)		MATERNAL (*n* = 44)		MERINO (*n* = 214)	
PC1 (25.3%)		PC1 (31.3%)		PC1 (25.6%)	
Metric	**Weight**	**Metric**	**Weight**	**Metric**	**Weight**
ageped	-0.38	avpedknown	0.35	avpedknown	0.38
avpedknown	0.36	ageped	-0.34	fullped	0.38
time_weights	-0.35	fullped	0.31	ages	0.36
time_carcase	-0.30	link_repro	0.28	bt	0.34
fullped	0.27	eff	0.27	dayinweek	-0.31
eff	0.26	time_carcase	-0.26	ageped	-0.29
link_weights	0.26	dayinweek	-0.25	rec_repro	0.27
bt	0.26	time_weights	-0.25	eff	0.26
dayinweek	-0.22	rec_weights	0.23	link_weights	0.20
link_repro	0.22	link_wool	0.22	link_wool	0.19
cnt	0.21	link_carcase	0.21	time_weights	-0.15
ngeno	0.18	link_weights	0.19	time_wool	-0.13
rec_weights	0.17	ages	0.19	synped	-0.13
ages	0.12	time_repro	-0.18	ngeno	0.12
link_wool	0.08	synped	-0.17	link_carcase	0.08
rec_wool	0.07	ngeno	0.13	rec_wool	0.03
synped	-0.03	rec_repro	0.13	rec_weights	0.01
		time_wool	-0.11		
		rec_wool	0.01		

The amount and depth of pedigree (**fullped** and **avpedknown**) were given the highest weightings across all breeds. In general, the direction of the weights were as expected – positive weightings, and hence higher DQS scores, were allocated for more pedigree, effective progeny numbers, increased linkage and variation in date of births. Negative weightings, resulting in lower DQS scores, were allocated for less timely submission of pedigree and trait data, more syndicate pedigree, and less variation in DOB days of the week. Therefore, capturing the variation in data characteristic metrics through PC1 provide reasonable weighting estimates to capture flocks with good data characteristic metrics (i.e. high DQS scoring flocks).

Use of the second principal component (PC2) was explored, as PC2 explained a further 11% to 14%. The emphasis placed on metrics for PC2 were in the opposite direction of PC1, which is expected since PC1 and PC2 were constructed as independent, orthogonal variables. Therefore, use of PC2 weightings capture the flocks with suboptimal data characteristic metrics (i.e. low DQS scoring flocks). The simultaneous use of both PC1 and PC2 weightings, further weighted by the proportion of variance explained, was also explored but yielded the same results as use of PC1 only (results not shown).

Method 2: Stepwise MLR (DQS_indexacc, DQS_avindex, DQS_indextrend, DQS_GG)

Index accuracy, average index and index trend were used to 'train' the DQS. Since there are benefits and limitations for use of these metrics, a new 'gains' metric was derived to capture all aspects of index accuracy, average index and index trend (Table 9), named GG_PC1. All genetic gains metrics were fairly equally weighted in the overall genetic gains metric. This also resulted in GG_PC1 being strongly correlated with all gains metrics, with Pearson's correlations ranging from 0.80 to 0.86 for in Terminal flocks, 0.74 – 0.85 for Maternal flocks and 0.67 – 0.81 in Merino flocks.

Table 9. An overall genetic gains metric combining index accuracy, average index and index trend through the first principal component (PC1)

TERMINAL (n = 396)		MATERNAL (n = 91)		MERINO (n = 256)	
PC1 (69.8%)		PC1 (61.2%)		PC1 (51%)	
Metric	**Weight**	**Metric**	**Weight**	**Metric**	**Weight**
Indexacc	0.60	Indexacc	0.53	Indexacc	0.66
Avindex	0.58	Avindex	0.59	Avindex	0.52
Indextrend	0.55	Indextrend	0.61	Indextrend	0.55

Stepwise regression analysis of **indexacc** and genetic gains metrics was conducted for metric reduction and estimation of weights for data characteristic metrics (Table 10). Model fit was assessed by examining residual plots. Between 23% to 86% of variation in **indexacc** and genetic gains metrics was explained, with **indexacc** explaining the most across all breed types (range 80 – 86%). This is due to **indexacc** being the metric most independent of breeder selection decisions, which was not accounted for in this analysis.

In general, the direction of the weights were also as expected, except for some metrics (highlighted yellow in Table 10). Again, this may be a reflection of genetic gains being influenced by factors beyond data characteristic, in particular selection decisions. Training against **indexacc** provide the most reasonable weighting estimates in the right direction. These results allow decomposition of **indexacc**, which can provide a basis for informing ram breeders for how data recording can be improved.

Table 10. Partial regression coefficients and coefficient of determination (R²adj) from stepwise regression analysis of index accuracy, average index, index trend, and all 3 combined (GG_PC1). Metrics in red text have partial regression coefficients that are not in the expected direction

Breed	Indexacc (R^2_{adj} = 0.80)			Avindex (R^2_{adj} = 0.47)			Indextrend (R^2_{adj} = 0.38)			GG_PC1 (R^2_{adj} = 0.70)		
TERM	**Metric**	$\hat{\beta}$	**SE**	**Metric**	$\hat{\beta}$	**SE**	**Metric**	$\hat{\beta}$	**SE**	**Metric**	$\hat{\beta}$	**SE**
(*n* = 354 flocks)	avpedknown	3.38	0.25	avpedknown	3.60	0.45	link_weights	0.36	0.06	avpedknown	0.67	0.05
	link_weights	1.70	0.18	link_repro	1.41	0.32	link_repro	0.24	0.07	link_weights	0.38	0.05
	ngeno	1.82	0.25	ngeno	1.83	0.53	ngeno	0.28	0.09	ngeno	0.43	0.07
	rec_weights	0.91	0.17	ages	-1.19	0.39	rec_wool	-0.16	0.07	link_repro	0.21	0.04
	ages	0.73	0.21	fullped	-1.15	0.45	time_weights	-0.12	0.06	rec_weights	0.15	0.04
	dayinweek	0.65	0.22	dayinweek	-1.02	0.44	fullped	0.14	0.07	time_weights	-0.14	0.05
	ageped	-0.54	0.20	link_weights	0.83	0.37	rec_weights	0.10	0.05	rec_wool	-0.09	0.04
	cnt	0.35	0.15	time_weights	-0.75	0.35	avpedknown	0.12	0.08			
	fullped	0.36	0.22	rec_wool	-0.62	0.29	link_wool	0.10	0.07			
	M.F	0.51	0.32				rec_repro	-0.10	0.07			
	Indexacc (R^2_{adj} = 0.86)			**Avindex (R^2_{adj} = 0.43)**			**Indextrend (R^2_{adj} = 0.39)**			**GG_PC1 (R^2_{adj} = 0.61)**		
MATL	**Metric**	$\hat{\beta}$	**SE**	**Metric**	$\hat{\beta}$	**SE**	**Metric**	$\hat{\beta}$	**SE**	**Metric**	$\hat{\beta}$	**SE**
(*n* = 70 flocks)	fullped	3.02	0.46	link_weights	6.86	1.46	time_weights	-0.31	0.12	link_weights	0.52	0.14
	rec_weights	1.07	0.30	fullped	4.95	1.53	link_repro	0.34	0.13	cnt	0.30	0.10
	rec_repro	0.92	0.32	cnt	2.93	1.03	fullped	-0.38	0.15	rec_weights	0.33	0.11
	M.F	0.74	0.26	link_wool	-2.63	1.25	rec_wool	-0.24	0.11	link_wool	-0.26	0.11
	link_weights	1.11	0.44	eff	-2.80	1.44	rec_weights	0.23	0.12	ngeno	-0.22	0.11
	dayinweek	1.07	0.46	bt	-2.42	1.47	cnt	0.20	0.10	synped	-0.19	0.09
	eff	0.76	0.35	rec_repro	2.07	1.28	synped	-0.15	0.11	time_weights	-0.21	0.11
	link_repro	0.71	0.36	ngeno	-1.50	1.11				link_repro	0.28	0.15
	ageped	-0.69	0.43							rec_repro	0.20	0.13
	time_repro	-0.47	0.30									
	cnt	0.34	0.24									
	Indexacc (R^2_{adj} = 0.83)			**Avindex (R^2_{adj} = 0.23)**			**Indextrend (R^2_{adj} = 0.26)**			**GG_PC1 (R^2_{adj} = 0.61)**		
MERI	**Metric**	$\hat{\beta}$	**SE**	**Metric**	$\hat{\beta}$	**SE**	**Metric**	$\hat{\beta}$	**SE**	**Metric**	$\hat{\beta}$	**SE**
(*n* = 214 flocks)	avpedknown	2.52	0.38	link_wool	4.10	1.14	ngeno	0.45	0.13	ngeno	0.25	0.06
	rec_wool	1.57	0.26	rec_wool	2.83	1.02	ageped	0.57	0.17	rec_wool	0.25	0.06
	link_wool	1.14	0.29	ngeno	2.26	0.97	link_repro	0.48	0.15	avpedknown	0.31	0.09
	fullped	1.66	0.42	ageped	2.40	1.03	link_weights	0.72	0.22	link_repro	0.28	0.09
	rec_repro	1.29	0.36	link_repro	2.45	1.07	fullped	0.42	0.15	ages	-0.17	0.07
	ageped	-0.81	0.28	ages	-1.88	1.04	time_weights	-0.38	0.15	fullped	0.22	0.10
	link_repro	1.08	0.39	cnt	1.65	0.93				ageped	0.15	0.07
	ngeno	0.63	0.24	M.F	1.50	0.87				link_wool	0.20	0.12
	eff	0.82	0.33							link_weights	0.26	0.18
	synped	-0.65	0.26							rec_repro	0.13	0.09
	ages	-0.72	0.30							time_weights	-0.09	0.06
	rec_weights	0.79	0.33									

Method 3: Informed weightings

This weighting structure was derived using taking into account the previous methods outlined, as well as understanding variation in each metric, relationships between metrics, and the breeding objectives of each breed type (Table 11). More weighting was placed on metrics that required the most improvement. For example, the amount of full pedigree was notably lower for MERI flocks compared to TERM and MATL flocks. Therefore, a higher weight was placed on this metric for MERI flocks.

Table 11. The data quantity, quality and timeliness metrics, and their relative weightings, contributing to the proposed Data Quality Score.

< to remain confidential>

DQS Calculation

The metrics and their weightings derived from the 3 methods described were used to calculate an overall DQS:

$$DQS = Weight_1 \times Metric_1 + Weight_2 \times Metric_2 + \ldots + Weight_n \times Metric_n$$

Where DQS is the data quality score, $weight_n$ is the weight for scaled metric n.

For ease of interpretation, the Data Quality Scores were then further rescaled so that scores ranged from 0 to 100 (Figure 16). (Note that scores were later scaled to range between 1 and 99 during implementation phase).

Figure 16. Distributions of a) raw and b) scaled Data Quality Score (DQS) scores, derived from principal component analysis (DQS_PCA), trained on index accuracy, average index, index trend, a combination of all three (DQS_indexacc, DQS_avindex, DQS_indextrend, DQS_GG_PC1) and informed weightings (DQS_Informed)
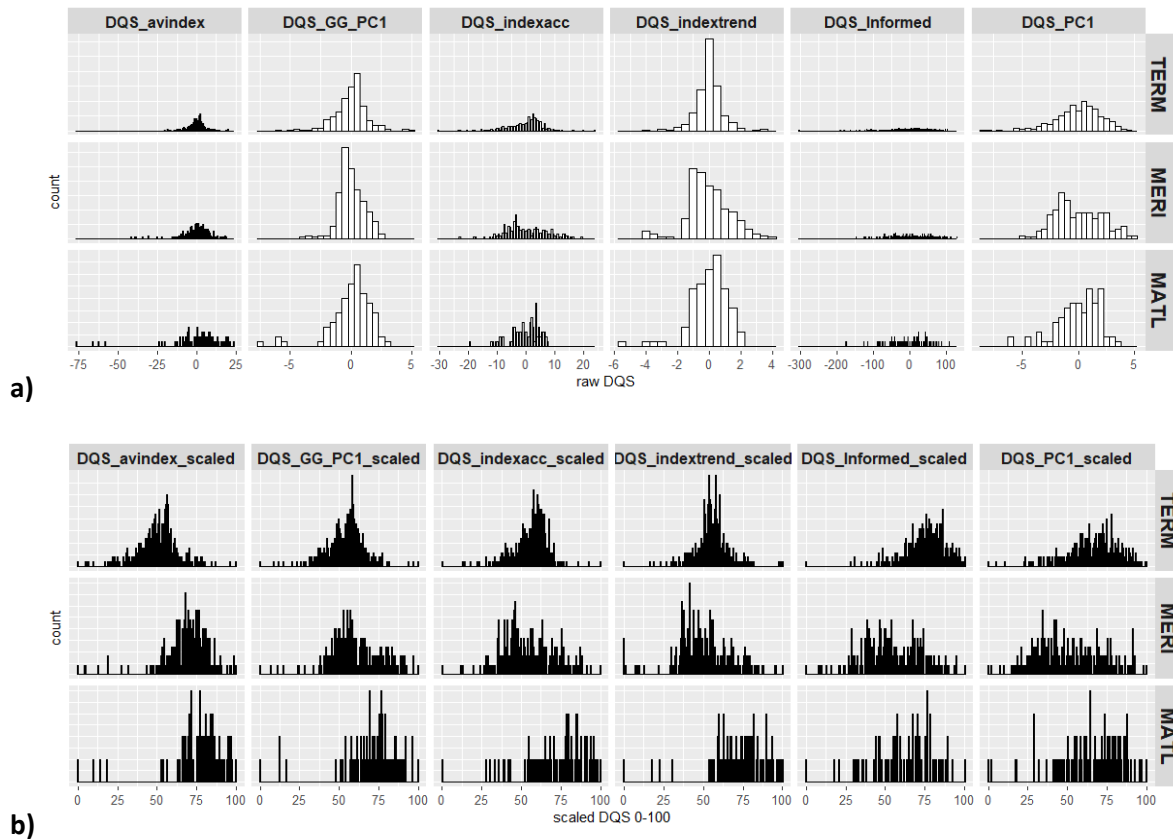
The correlations between the scaled DQS scores were generally high (Figure 17), especially for Terminal flocks (range of 0.75 to 0.97). The correlations were not as strong for Merino and Maternal flocks but were generally still high except for indexes derived from average index and index trend (DQS_avindex, DQS_indextrend). This may be due to the varying breeding objectives and indexes used in Merino and Maternal Flocks. If disregarding the DQSs derived from average index and index trend, Pearson's correlations between the remaining indexes range from 0.86 to 0.93 for Maternal flocks, and 0.82 to 0.96 for Merino flocks. Therefore, there is limited re-ranking of flocks for the DQSs are derived independent of metrics that describe genetic gains (based on specific index calculations).

The DQS_PC1 and DQS_Informed scores were derived independently of, but related to, metrics that describe genetic gains. Pearson's correlations between these 2 DQS methods and genetic gains metrics ranged from 0.5-0.81 for Terminal flocks, 0.11 to 0.80 for Merino flocks and 0.4 to 0.86 for Maternal flocks (Figure 17). This highlights that flocks with higher DQSs also made greater genetic gains, which demonstrates value for an overall DQS score. This will be further highlighted in Section 4, when recommended format and scales are provided.

Figure 17. Pearson's correlations between index accuracy (indexacc), average index (avindex) and index trend (indextrend) and Data Quality Score (DQS) scores, derived from principal component analysis (DQS_PCA), trained on index accuracy, average index, index trend, a combination of all three (DQS_indexacc, DQS_avindex, DQS_indextrend, DQS_GG_PC1) and informed weightings (DQS_Informed). Calculated for 396 Terminal (TERM), 256 Merino (MERI) and 91 Maternal (MATL) flocks
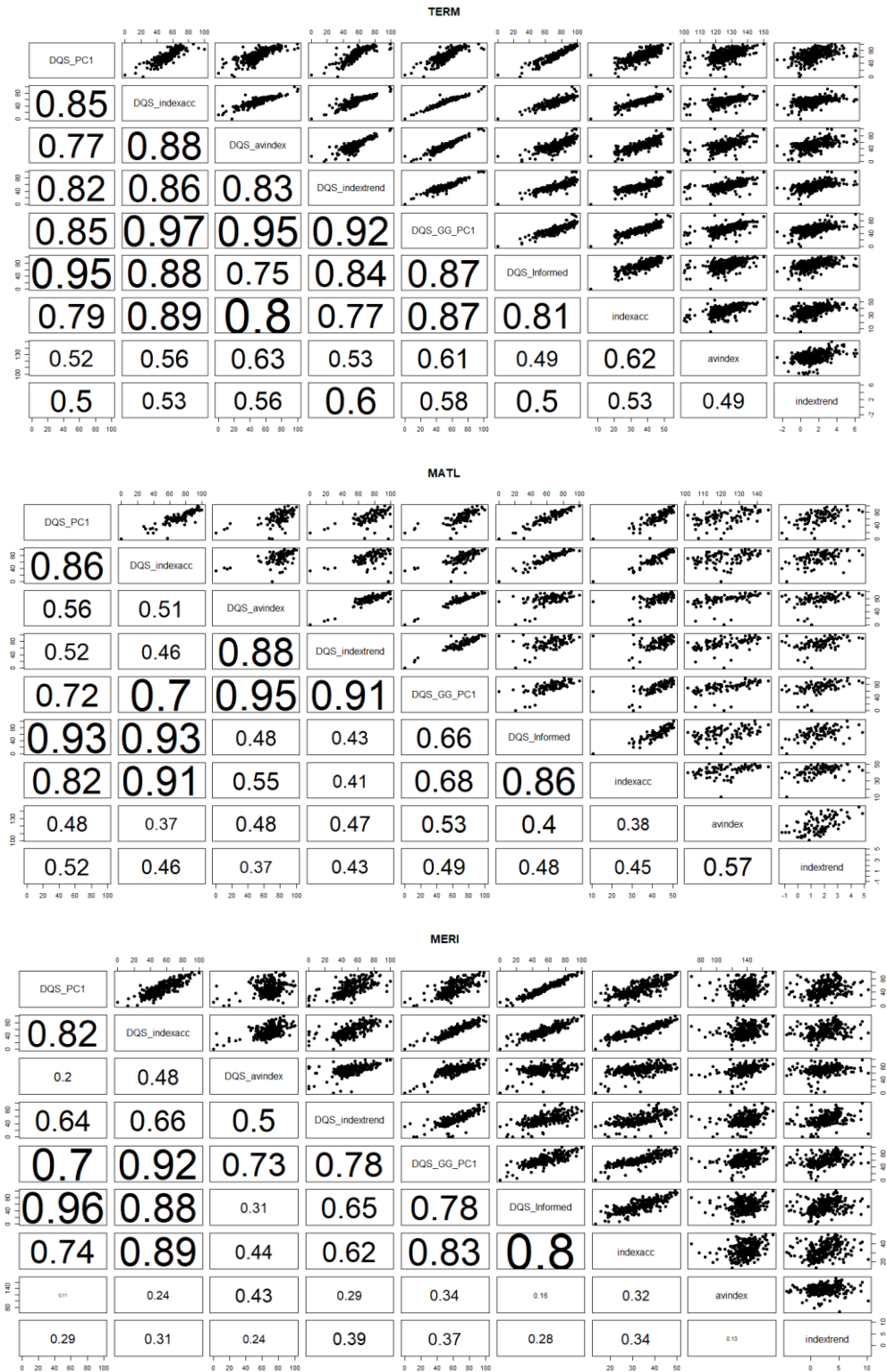
The DQS was also provided to Sheep Genetics and examined to ensure they met expectations.

There was no definitive or optimal method for how an overall DQS is calculated. Each method has its benefits and limitations:

- o The PCA approach examines variation in data quality metrics alone. However, the principal component weightings depend on what other variables are included in the analysis.
- o The MLR approach to train the DQS against genetic gains metrics is not optimal because genetic gains can be attributed to factors apart from data recording (for example, breeder decisions on use of outside genetics, selection differential etc.). The approach of training on genetic gains can be viewed undesirably by industry as it favours flocks with higher genetic gains, and also assumes that breeders are using the indexes examined.
- o The informed weightings approach can be viewed as subjective, albeit based on research findings.

For all approaches explored, flocks with higher DQS scores had higher index accuracies (r = 0.86 for MATL, r = 0.81 for TERM and r = 0.80 in MERI flocks). Index accuracy can be seen as the most appropriate current metric to compare the DQS to, as it is independent of breeder selection decisions and is a measure that reflects the data used to estimate breeding values. Nevertheless, flocks that had better quality data (quantified through either DQS and index accuracy) also had higher genetic gains (avindex and indextrend).

The recommendation for the first prototype of the DQS is the informed weightings approach. The calculation of this DQS is independent of, but related to, genetic gains. This approach provides a balance between scientific rigour with industry adoption, and is easily understood and interpreted. Road-testing with industry stakeholders will then assist in further refinement. From here on, this report will discuss the DQS derived through the informed weightings approach.

4.3 Demonstration of data quality metric reporting

Data Quality Score – Star ratings

Due to the informed albeit subjective nature of the weightings used to calculate the DQS, the DQS was categorised as 'star ratings'. It is envisioned that these scores and star ratings will initially will be reported privately in RUGG reports, along with metrics as specific breeder feedback. There is support to publish star ratings publically after a period of road-testing with industry.

Three methods were explored for partitioning of the DQS: a) quintiles (equal number of flocks per star rating), b) equal split in DQS range (0-20, 20-40, 40-60, 60-80,80-100) and c) normal distribution in percentiles – 10%, 20%, 40%, 20%, 10% in star ratings 1 to 5, respectively. The distributions of these star ratings are provided in Figure 18.

Figure 18. Data Quality Score (DQS) overall score, derived through the informed weightings approach, partitioned into star ratings using a) quintiles, b) equal range, and c) normal distribution percentiles
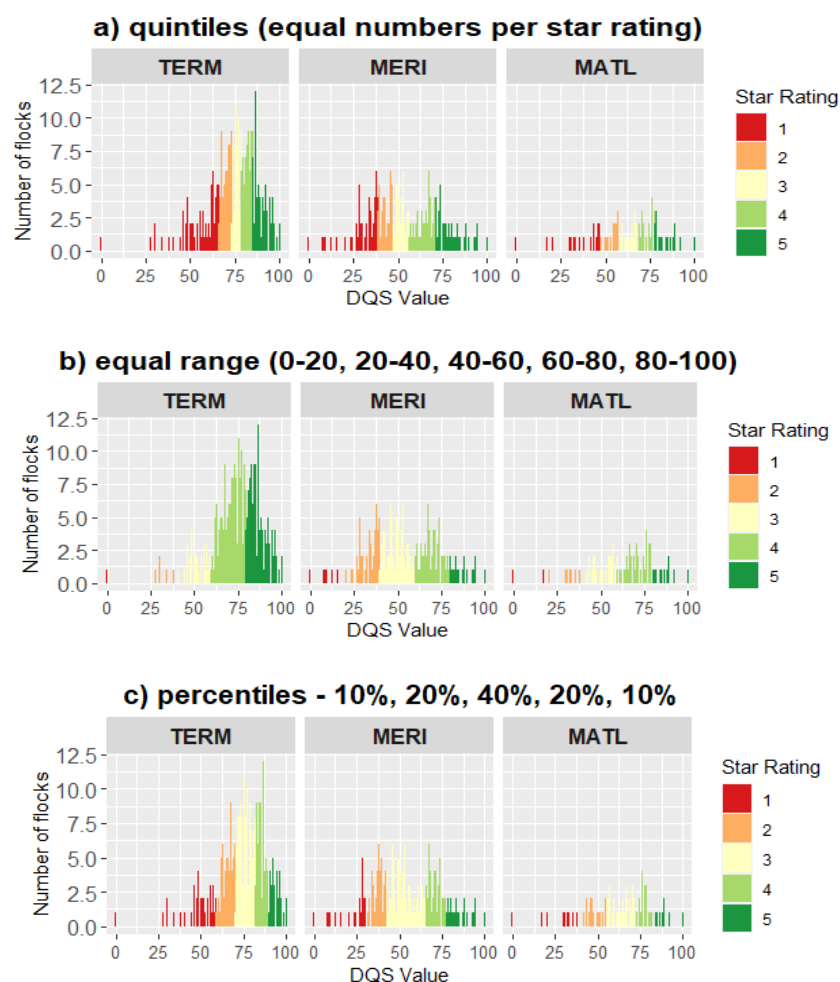
The relationships between genetic gains and DQS Star ratings were explored. When partitioning DQS scores by quintiles (option a), the overlapping distributions for each star rating indicate that there are no significant differences in **indexacc**, **avindex** and **indextrend** across the star ratings. The differences between extreme star ratings are more prominent and significant using equal range (option b) and

normal distribution of flocks (option c). Options a and c also resulted in more flocks with 1 star rating compared to partitioning by equal ranges (20% and 10% flocks vs. 0.3-2.3% flocks). Therefore, the recommendation is to partition DQS scores into star ratings through equal range (option b).

4.3.1 Data Quality Framework Reporting

Since the proposed DQS score has used metrics already derived in RUGG reports, it is recommended that the overall DQS score, ratings, and refined metrics calculated be incorporated into RUGG report.

We propose that the RUGG includes the following aspects from this project:

- The Data Quality Score
- Star rating
- Where they sit compared to other flocks
- Highlight data recording strengths
- Provide recommendations for where recording can be improved
- New/refined metrics – e.g. recorded and linkage by trait group, day in week, timeliness of data. These can be incorporated into existing categories (e.g. 'Pedigree Analysis') or new categories (e.g. "Timeliness")

New software was written to generate interim DQS reports to demonstrate how it can be incorporated into RAMping Up Genetic Gains reports. Note that Version 2 of the RUGG report has been used.

RAMping Up Genetic Gain

MOCK UP report to include Data Quality Score prototype

RAMping Up Genetic Gain reports aim to provide breeders with better information on the variables that impact the rate of genetic gain of their flock. These variables include the amount and quality of pedigree, the quality of the data and how effectively this data is used. By considering the individual components of the breeders equation and balancing them breeders are able to maximise genetic gain.

The RAMping Up Genetic Gain reports help evaluate your flock's breeders equation. RAMping Up Genetic Gain reports are a tool breeders can use to summarise data from their flock into these components that influence genetic gain.

The context information supplied below each report section provides a benchmark of how your flock is performing relative to the rest of the analysis for factors impacting genetic gain. Context information includes:

- Flock Average - the five year average of the flock the report is based on for the component in question
- Analysis Average - the five year analysis average for the component in question
- Rating - a five-star rating of how your flock compares to the rest of the analysis for that component of genetic gain. For example a five-star rating means that your flock is in the top 20% of the analysis for that component. A one-star rating means your flock is in the bottom 80% for that component.

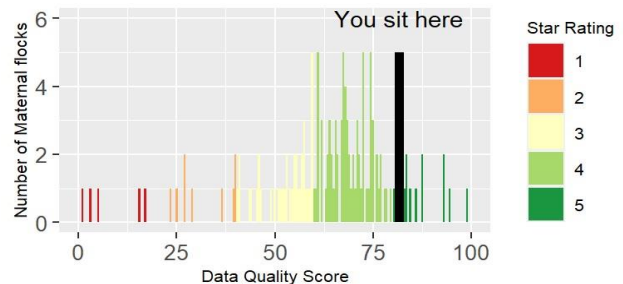**Flock Code:** GENERIC. **Site Code:** GENERIC (MATL flock). **Flock Name:** GENERIC

Date Created: 08 June 2021 from run 17May21.

Data Quality Score: 81.7

★ ★ ★ ★ ★

5 star

Your Data Quality Score (DQS) is an overall indicator of the quantity, quality and timeliness of your data.

Your data recording strengths:
- Depth of pedigree
- Recording full pedigree
- Recording of birth dates

To improve your data quality:
- Record more animals for reproduction traits
- Record more animals for weight traits
- Submit reproduction trait data earlier

Although key recommendations have been provided, refer to the entire RUGG report to decide which components of data recording are most relevant to your flock.

Pedigree Analysis

Date of birth recording
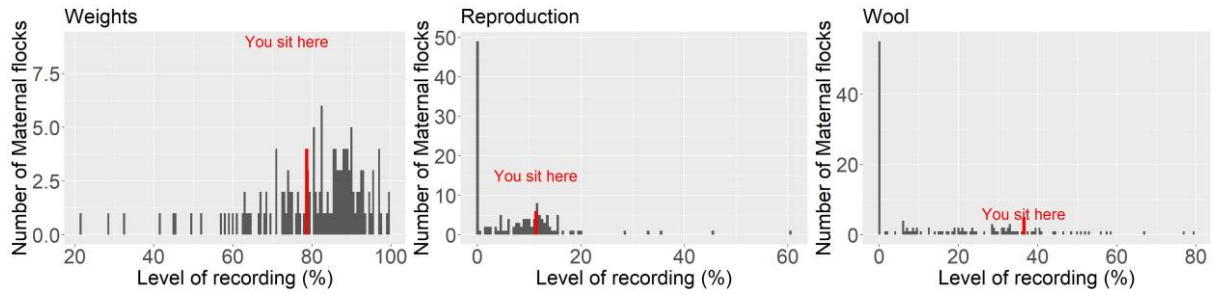
Number of unique birth dates

Year	2015	2016	2017	2018	2019	2020
Unique birth dates	53	60	70	51	53	48

Data Summaries

Average recording by trait groups

Weights	Reproduction	Wool
78.6 %	11.3 %	36.6%

Average linkage by trait groups

Weights	Reproduction	Wool	Carcase
62.8%	48.6%	43.5%	58.4%
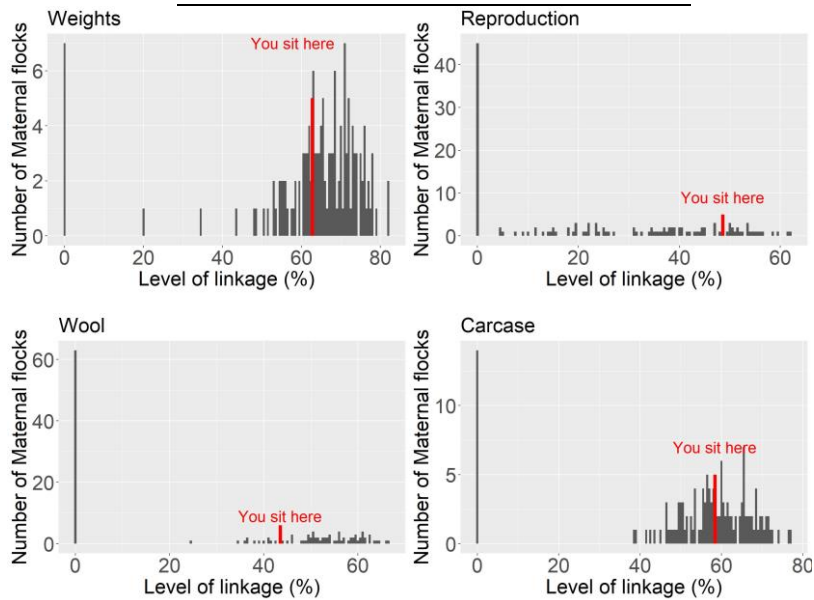
Timeliness

Pedigree Data

Average time to submit pedigree data

Year	2016	2017	2018	2019
Average age (days)	137	330	208	73

Overall average 117days

5 Stars
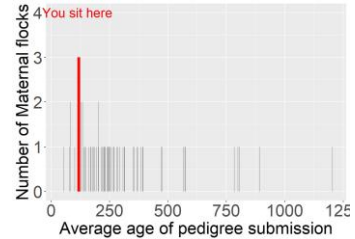
Trait Data

Average time to submit trait data after measurement

	Weight	Reproduction	Wool	Carcase
	54 days	67 days	111 days	16 days
	4 Stars	3 Stars	1 Star	4 Stars
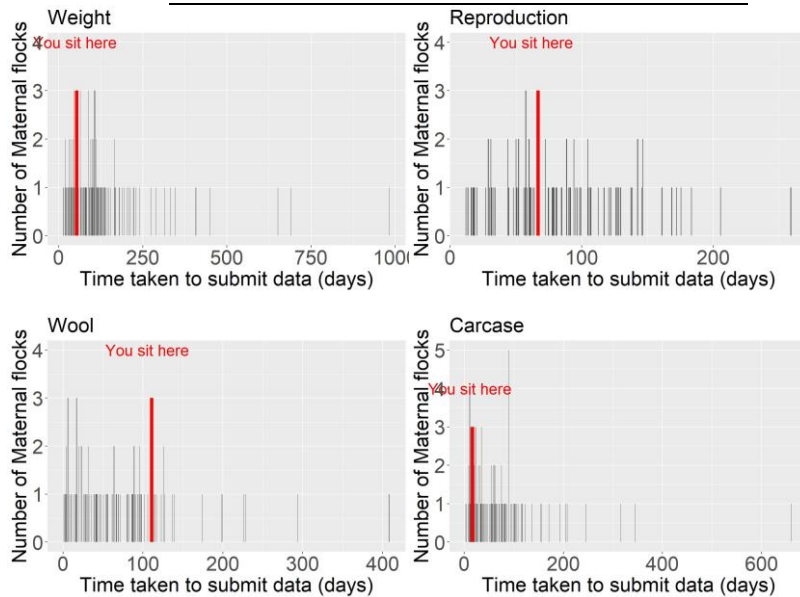
4.3.2 Feedback and road-testing with industry

The methodology used to derive the data quality framework and DQS was discussed in various forums:

- **AGBU Seminar** – 15th February 2021
- **BREEDPLAN and Sheep Genetics Technical Committee** – 3rd March 2021

The concept of the data quality framework and methodology explored to derive the overall data quality score was outlined. Endorsement was given to progress to implementation and road-testing of the first prototype of the DQS (based on the informed weightings method).

Considerations highlighted included:

- *Cost benefit analysis*
  - To what extent should recording be improved, considering the costs associated?
- *Value proposition*
  - How can we better demonstrate the value proposition and implications of a DQS ? Potentially a simulation to understand how changes in data recording reflect consequent outcomes in genetic gains
- *Fixed effects*
  - How do we better capture completeness and accuracy of fixed effect recording?
- *DQS reference vs. DQS individual breeder*
  - The proposed DQS characterises the data, with the purpose of understanding the value of the data to individual breeders and their clients. How can this be applied to valuing data contributing to the reference population (or other breeders)? These DQSs will be related but capturing different aspects of data value.

- **Sheep Genetics Technical Operations meetings**

Continuous communication with Sheep Genetics was maintained to ensure efficient implementation of the data quality framework calculations in OVIS. At the time of this report, funding was not available to incorporate reporting of the DQS prototype and associated metrics/reporting into RUGG reports. It was decided an interim report will be sufficient in the initial road-testing phase.

- **Internal AGBU meetings** - multiple meetings to discuss the project scope, metrics used and methodology used to derive the DQS
- **Conference papers:** Association for the Advancement of Animal Breeding and Genetics (AAABG) conference – 2 papers submitted (see Appendix 2)

ROAD-TESTING AT INDUSTRY EVENTS

The data quality framework was road-tested with industry at 6 separate events, with interim DQS reports generated for 96 flocks.

Communication approach

- The specific metrics used to calculate the DQS were not disclosed (although the metrics are, or are anticipated to be, reported in RUGG reports). Data quality was described in reference to the data requirements for unbiased comparisons of animals, in particular the amount, the completeness and accuracy, the structure, and the timeliness of data. Figure 19 is the graphic used to outline the metrics used to calculate the score.

Figure 19. Defining data quality for the calculation of the Data Quality Score.


- The weightings the metrics used to calculate the DQS was also not disclosed. This was the approach taken to discourage 'cheating' or targeting improvement of only one metric. The description provided was
  "All measures are roughly equal, with some adjustments depending on:
  - *Breed type*
  - *Level of trait recording*
  - *Variation and relationships between characteristics"*

- The first workshop with service providers highlighted the potential confusion between ASBV accuracy and the DQS. Future workshops required explanation of the difference between the two (Figure 20).

Figure 20. The difference between ASBV accuracies and the new Data Quality Score
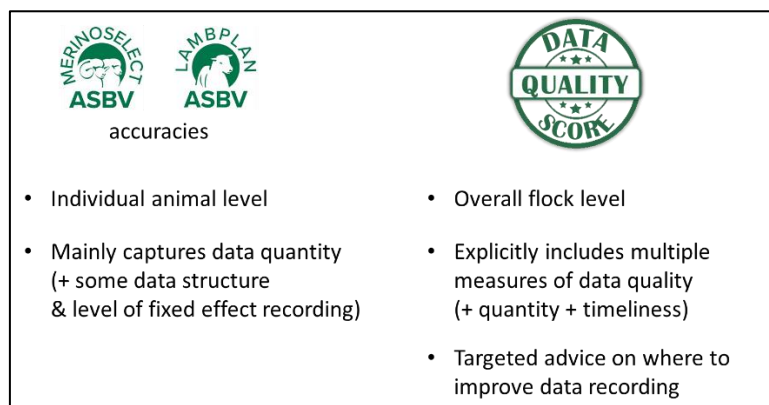
Industry response to the DQS

The concept of the DQS and its reporting through RUGG reports was well received and feedback was positive. There were questions regarding the practicalities of improving data quality, the cost benefit, and the value of some metrics. These questions will require further investigation to be answered.

Feedback was provided through live anonymous polling during workshop sessions to stimulate group discussion, as well as anonymous feedback sheets provided at the end of the session. Note that not all participants provided feedback, and responses are not available for all questions due to questions being refined over time with subsequent events. This report will only discuss feedback from the written feedback sheets (except a subset of questions for the Launceston Regional Forum, where responses were from live anonymous polling). Table 12 shows the number of responses for each event, number of flocks for which DQS reports were generated, and the average and range in DQS scores for the flocks at each event.

Table 12. Data Quality Score (DQS) report road-testing: number of feedback responses, number of flocks for DQS reports were generated and their average DQS scores across 7 events

Event	Service Provider workshop	Armidale Regional Forum	Launceston Regional Forum	Adelaide Regional Forum	Dubbo Regional Forum	MerinoLink	$uperborder$ conference	Total
Number of feedback responses	13	7	5	8	8	11	11	**63**
Number of DQS reports generated (flocks)	14	16	8	16	11	31	14	**101**
Average DQS score [range]	65 [26 - 92]	69 [44 - 92]	77 [68 - 85]	76 [38 - 99]	68 [35 - 99]	66 [39 - 99]	69 [52 - 88]	**69 [25 – 99]**

There were 101 flocks represented across the 8 events, and feedback responses were received from 63 participants. The average DQS score (across all breed types) was 69, which ranged from 25 to 99. Flocks represented at the Service Provider workshop had the most diverse data quality (standard deviation of DQS scores of 21.7).

Use of the DQS to industry

The majority of attendees (88% overall, 55 out of 63 respondents) thought that the DQS and star rating will be useful and relevant for industry (Figure 21). There were 2 respondents who disagreed in the Service Provider workshop, and 1 strongly disagreed from the Dubbo Regional Forum. There were 4 respondents who were uncertain about the use of the DQS.
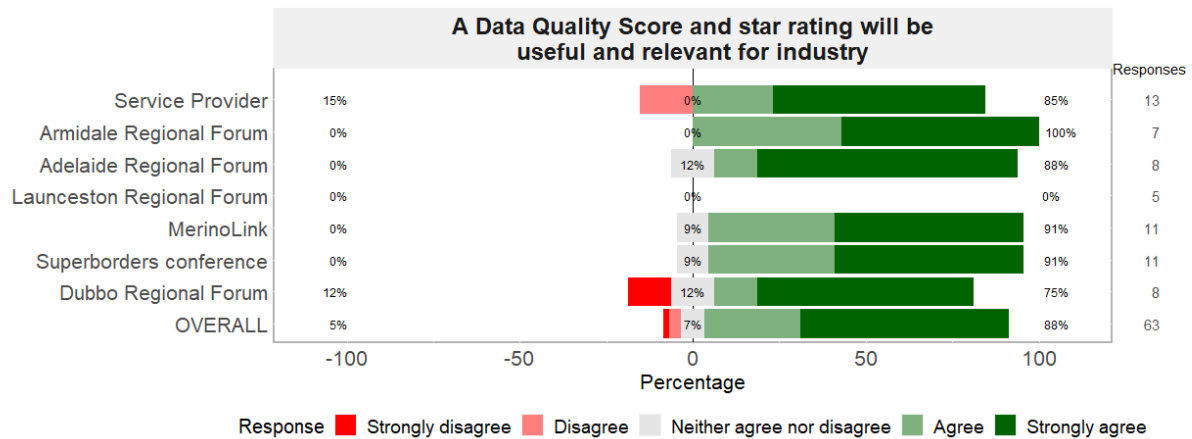

Figure 21. Level of agreement (green) and disgreement (red) for the usefulness and relevance of the Data Quality Score (63 survey responses)

Breeders were asked how the DQS and star rating would be used by industry. Responses are summarised in the wordcloud in Figure 22. From the 35 respondents, the words written more than 5 times were 'stud', 'quality', 'improving', 'ram', 'accurate', 'buying', and 'confidence'.
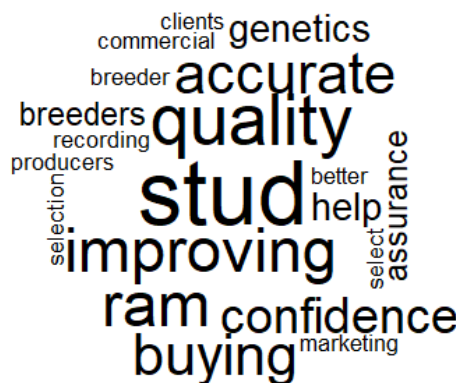

Figure 22. Wordcloud for responses to the question "How do you see the Data Quality Score and star rating be used by industry?" (43 responses)

Usefulness of DQS as a tool for Service Providers

The RUGG report with the DQS metrics (which was previously called Data Quality Index, DQI) was seen to be a useful tool for 85% (11 out of 13) of service providers for giving advice to clients (Figure 23). Service Providers are an important support service and conduit to Sheep Genetics breeders, and so the DQS report is an important tool for providing targeted advice on data collection and management.
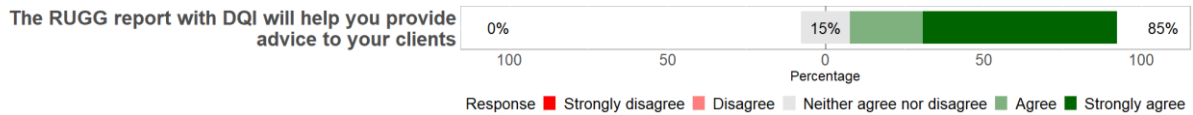
Figure 23. Level of agreement (green) and disgreement (red) for the usefulness of RAMping Up Genetic Gain (RUGG) reports updated with Data Quality Score metrics (formally Data Quality Index, DQI), for service providers to provide advice to clients (13 responses)

Usefulness of DQS for breeders to improve data recording

There were 50 breeders that responded to the question regarding DQS enticing change in their data recording, with 44 (88% overall) stating that there will be lots or some change (Figure 24). There were 3 breeders (from the Launceston Regional Forum and Dubbo Regional Forum) who thought there will not be any change in their data recording. Note that the feedback from Launceston Regional Forum was supplied through the live slido polling during the workshop.
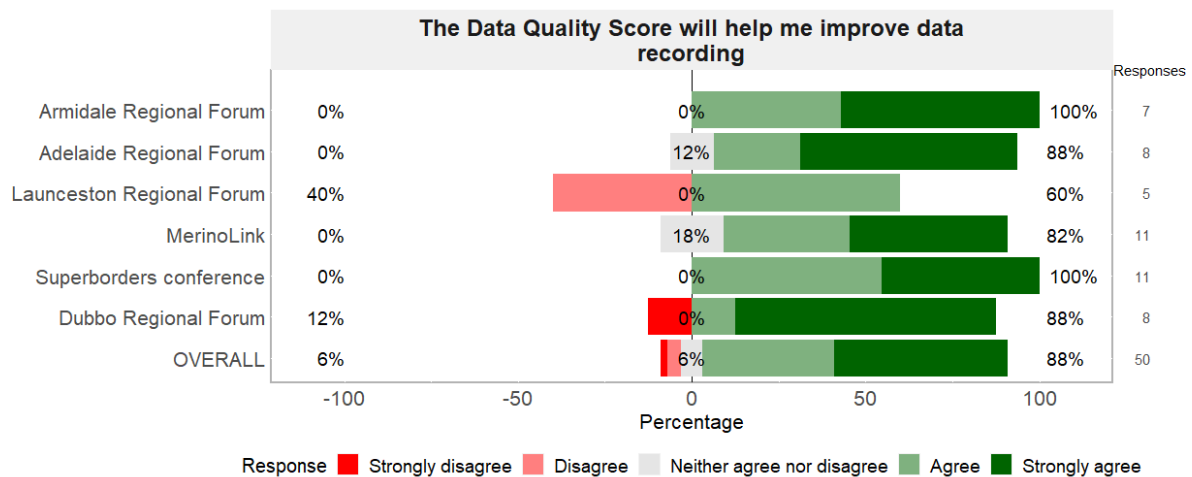

Figure 24. Level of agreement (green) and disgreement (red) for the usefulness of Data Quality Score for breeders to improve data recording

Breeders were asked how if the DQS would help them think differently about data. Responses are summarised in the wordcloud in Figure 25. From the 43 respondents, the words written more than 5 times were 'yes', 'improve', 'timeliness', and 'accuracy'.

Figure 25. Wordcloud for responses to the question "Will the Data Quality Score help you think differently about data quality? If so, how?" (43 responses)

Public disclosure of the Data Quality Score and Star rating

The majority of breeders (43 out of 48 respondents, 90%) were in favour of public disclosure of the DQS and star rating, providing there was a grace period.

4.3.3   The value proposition for the Data Quality Score

The DQS was calculated independently of, but related to, index accuracy & genetic gains. Flocks that had higher data quality scores had significantly higher index accuracies and higher rates of genetic gains (Figure 26).
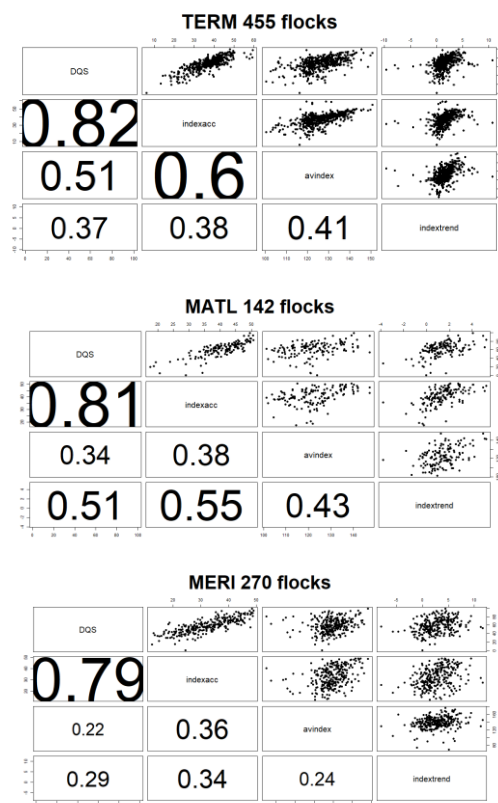

Figure 26. Pearson's correlations (below diagonal) and scatter plots (above diagonal) to demonstrate the relationships between the Data Quality Score (DQS) and index accuracy (indexacc), average index (avindex) and index trend (indextrend) in 455 Terminal, 142 Maternal and 270 Merino flocks.

The DQS accounted for between 63-67% of variation in **indexacc**; 5-26% in **avindex** and 8-26% in **indextrend**.

When examined on a DQS star rating level, there were significant differences in **indexacc** and genetic gains, particularly **indexacc** (Figure 27). This provides confidence to ram buyers that the DQS is a reflection of 1) the general quality of the flock they are buying from, and 2) this is also a reflection of the anticipated outcomes expected from the progeny of the rams they are going to buy.
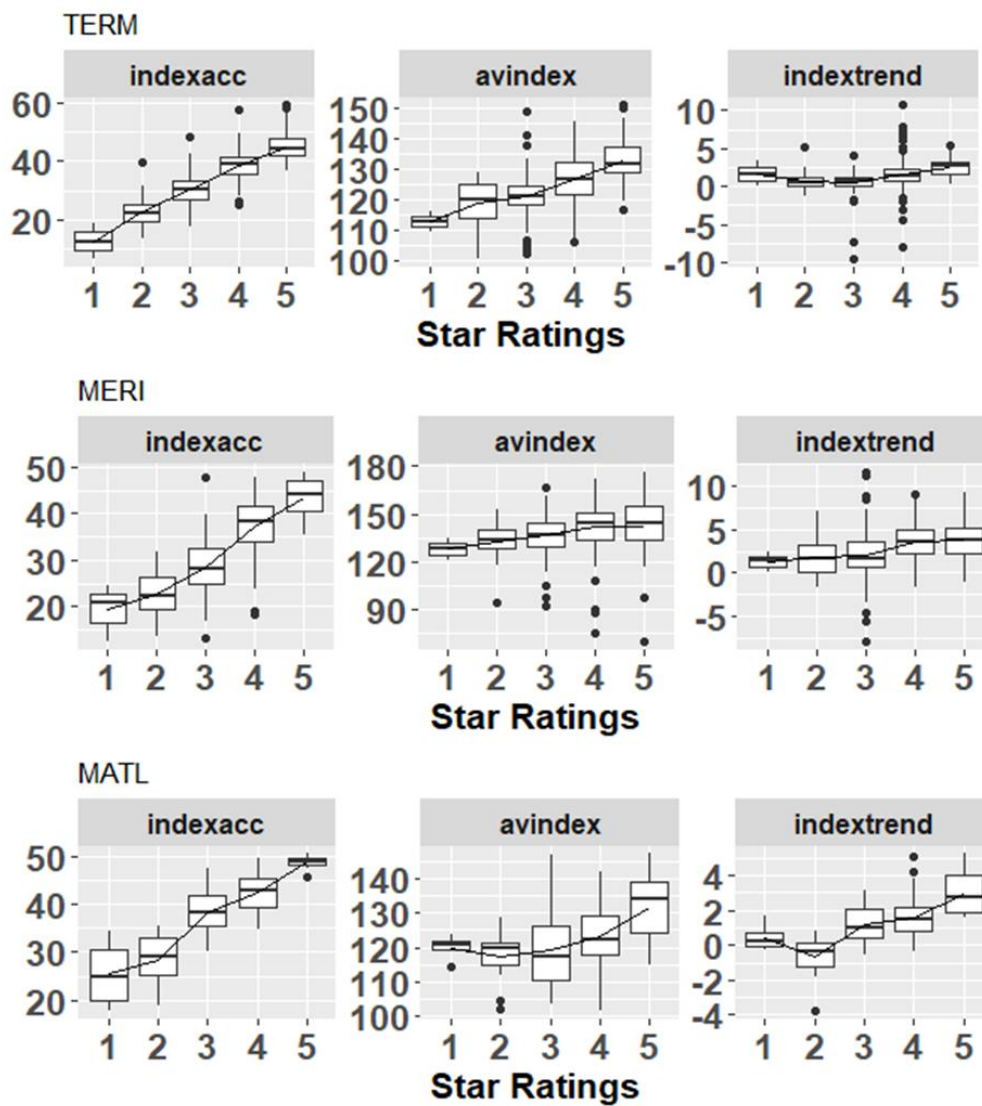
Figure 27. Higher Data Quality Score star rating flocks had higher index accuracies (indexacc), average index values (avindex) and index trends (indextrend) for Terminal (TERM), Merino (MERI) and Maternal (MATL) flocks

5. Software development for implementation in OVIS

There were 3 potential strategies for system implementation of the data quality framework:

1. Use of existing stoplights program
   New and refined metrics be incorporated into the existing stoplights program as subroutines. While this may be sufficient in this research phase, the process would be inefficient for routine analysis.
2. Re-write stoplights program
   This will allow faster processing of data characteristic metrics due to a more efficient program. However, this requires time and personnel.
3. Use existing DataAudit program
   Additional metrics were inspired by the metrics calculated in DataAudit. This option was explored as a way to potentially streamline the process for both beef and sheep.
   DataAudit was designed specifically for the BREEDPLAN analysis. Results are not provided automatically to each breeder or breed society, but is used as a diagnostic, research and engagement tool. Therefore, the output would not be suitable for breeders as feedback. Use of DataAudit in sheep will require extensive re-coding and changes in file structures. There is also uncertainty about how the use of DataAudit at Sheep Genetics fits into the ABRI commercialisation model of BREEDPLAN products.

Option 2 was taken in this project. The following steps were taken:

- **Creation of new timeliness database**
  Timeliness metrics were calculated for by iterating through every analysis run conducted in the last 5 years. Since this is a computationally demanding and time consuming process, a new timeliness database was created to store results. Results are updated with each new run.
- **Re-writing of stoplights program**
  This process involved
  - Updating the current stoplights program from Python 2 to Python 3
  - Calculation of timeliness data
  - Subroutine program to calculate DQS scores, star ratings, recommendations and strengths
  - Post-analysis saving of results as a pickle
    This has implications for the RUGG report API. This approach has significantly reduced the processing time (from 5 minutes to 20 seconds).

6. Conclusion

This project delivered a data quality framework to characterise the quantity and quality of genetic evaluation data, which has been implemented into the national sheep genetic evaluation system. Flocks with higher data quality scores achieved greater genetic gains and index accuracy. This work has been well-received by industry, and will assist ram breeders to target improvements in their data recording, provide recognition for their efforts, and support selection decisions to drive genetic gains.

6.1 Key findings

- There was substantial variation in the quantity, quality and timeliness of data across Terminal, Maternal and Merino flocks. This provides opportunities for improvement
- The Data Quality Score developed in this project was developed independent of, but is related to, metrics that describe genetic gains. Therefore flocks with better quality data made more genetic gains.
- Industry are positive about the implementation and usefulness of this data quality framework. Road-testing of the data quality framework with a small number of service providers and breeders show support for public publishing of the data quality score and star rating after a grace period.

6.2 Benefits to industry

- **An enhanced data feedback tool for breeders.**
  The RAMping Up Genetic Gains reports can be further enhanced with new data quality metrics, Data Quality Score, star rating, recommendations and strengths. This provides targeted advice to breeders to assist in management changes, improvements in data collection and submission and hence ASBV accuracy. In turn, this was assist in more accurate selection decisions and increased rates of genetic progress
- **Transparency for ram buyers** about the quality of data used to calculate EBVs. While EBV accuracies are available for individual rams, a DQS provides an indication of the overall quality of the flock's data and allows direct comparison across flocks.
- **A way to identify and highlight breeders** who collect high quality data. This could be used as a basis for discounted registration fees, or through breeder awards, or other signals and/or rewards
- **Engagement tool** for Sheep Genetics development officers (and service providers) for targeted extension activities for flocks with poor data quality.
- This data quality framework can be further developed to determine and value **data contribution to reference populations**

7. Future research and recommendations

The following research is recommended:

- **Understanding challenges and reasons for poor data recording**. Since some metrics are widely poorly recorded across many flocks (e.g. level of full pedigree recording in Merino flocks), it would be beneficial to understand 1) why it is poorly recorded, 2) explore/devise tools to increase ease of recording, and 3) design an extension campaign to target improvement of recording.

- **Cost-benefit analysis and tools** to understand to what extent it is worth improving data recording, considering the costs associated, at both the individual flock and the industry levels

- **Better demonstration of the value proposition.** Flocks with higher Data Quality Scores had higher index accuracies and rates of genetic gains. The Data Quality Score also provides additional information not captured in EBV/index accuracy.

  There is potential to undertake a simulation to better understand how changes in data recording reflect consequent outcomes in genetic gains.

  An important component over time will be to understand if RUGG and DQS reporting leads to change in behaviour and improved recording.

- **Better accounting of fixed effects.** Completeness and accuracy of fixed effect recording is only captured to a limited extent in the proposed framework. This requires more in-depth examination.

- **Data Quality Score -- reference vs. individual breeder.** The proposed DQS characterises the data, with the purpose of understanding the value of the data to individual breeders and their clients. Although related, an alternative perspective is valuing data contributing to the reference population, and/or other breeders.

PRACTICAL APPLICATION OF PROJECT INSIGHTS

- **Public and private reporting:** The recommended roll out strategy is to initially privately report the DQS whilst road-testing, before public reporting of star ratings after a grace period. The pathway to public release (including the length of the grace period) is yet to be fully defined.

- **Continued road-testing and education:** This is particularly important if there is a reward or incentive to having a high data quality score. This requires a detailed communication strategy, which may involve media releases, and fact sheets and videos on the Sheep Genetics website.

- **Incorporation into RUGG reports:** While an interim report is available, it would be ideal to incorporate the DQS and associated features into RUGG reports. Increasing the availability and use of the RUGG reports by service providers and breeders should also be a key strategy.

- **Continuous monitoring and refinement:** The metrics require monitoring, and weights require refinement over time. This will assist in evaluation of how effective the reporting is to entice change. There is also potential to further refine the DQS reporting

- **Understanding of poor data recording:** understanding recording challenges, and devising targeted extension messages

- **Application in beef:** The demand for an updated data quality framework for BREEDPLAN (currently delivered through DataAudit software) is unknown. In principle, frameworks for evaluating data quality should be consistent across species, primarily to facilitate extension, and potentially to simplify introduction of systems for valuing of data for reference populations.

DEVELOPMENT AND ADOPTION ACTIVITIES to ensure full value from project's findings

- **Continuous engagement with breeders:** The motivation to use the DQS can be encouraged with continuous communication, road-testing, requesting feedback from breeders about use and re-evaluation of reporting format.

  The RUGG report with DQS can be use as engagement tool at industry events or workshops e.g. Regional Forums.

- **DQS Resource:**

  Adoption will be enhanced if resources and information is easily accessible. This may be in the form of a page on data quality on the Sheep Genetics website, with direct links to tools that will help improve with data recording and the benefits to be seen resulting from the improvements.

- **Rewarding breeders:**

  An additional purpose of the data quality framework is to highlight the breeders that make a good effort. This could be used as a basis for discounted registration fees, or through breeder awards, or other signals and/or rewards.

  Alternatively poor scoring flocks may have EBVs withheld (which may already happen with low EBV accuracies).

A key contributor to success has been the collaborative nature of the project. There was consistent communication and input with MLA and Sheep Genetics from the project conception to DQS delivery. Constructive feedback from breeders has also been very beneficial for this project

8. References

Brown, DJ, Tier B, & Banks R 2001. Implications of data quality on the accuracy of estimated breeding values for weaning weight in sheep. Proceedings of the Association for the Advancement of Animal Breeding and Genetics, 14: 441-444

Brown, D.J., Swan, A.A., Gill, J.S., Balls, A.J., and Banks R.G. (2015). Genetic parameters for liveweight, wool and worm resistance traits in multi-breed Australian meat sheep. 1. Description of traits, fixed effects, variance components and their ratios. Animal Production Science, 56(9): 1442 – 14

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) 'ASReml User Guide Release 3.0' VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Huisman, A.E., Brown, D.J., Ball, A.J. and Graser, H.-U. (2008). Genetic parameters for bodyweight, wool, and disease resistance and reproduction traits in Merino sheep. 1. Description of traits, model comparison, variance components and their ratios. Australian Journal of Experimental Agriculture, 48: 1177 - 1185

James G, Witten D, Hastie T, Tibshirani R. (2017). "An Introduction to Statistical Learning: With Applications in R." 1st ed. 2013, Corr. 7th printing 2017 edition. Springer; 2013. https://faculty.marshall.usc.edu/gareth-james/ISL/ISLR%20Seventh%20Printing.pdf

Martin S. (2014). Valuing Data Quality. Final report (MLA project B.SGA.0182).

Stephen, L.M., Brown, D.J., Jones, C.E. and Collison, C.E. (2018). Key differences between high and low performing Australian Sheep flocks participating in genetic evaluation. Proceedings of the World Congress on Genetics Applied to Livestock Production, 11.433

Swan, AA & Brown, DJ 2007. A simple bootstrapping procedure to validate the MerinoSelect model for weaning weight. Proceedings of the Association for the Advancement of Animal Breeding and Genetics, 17: 396 – 398

9. Appendix 1 – Feedback forms

Service Provider: Data Quality Index (DQI) Workshop, May 2021

Small group discussion

Points for discussion

- What are your general thoughts about the DQI prototype? What is good? Is there anything that is missing?
- Will the DQI entice change?
- Does the DQI prototype reflect what you know about this flock's data?

Notes:

_____

_____

_____

Feedback on RUGG + Data Quality Index

Your input will help us further develop and enhance the usefulness of the DQI.

Please circle your level of agreement with the following statements	Strongly Disagree				Strongly Agree
A DQI score and star rating will be useful and relevant for industry	1	2	3	4	5
The RUGG report with DQI will help you provide advice to your clients	1	2	3	4	5
Reporting changes over time is important	1	2	3	4	5

1. Which component of data quality is most important to you and your clients?

_____

_____

2. How do you see the DQI score and star ratings being used?

_____

_____

3. How would you encourage your clients to think differently about data quality?

_____

_____

Any other comments?

_____

_____

Thank you!

agbu **heep** GENETICS mla

Feedback on the Data Quality Score Prototype

Your input will help us further develop and enhance the usefulness of the Data Quality Score.

Please circle your level of agreement with the following statements	Strongly Disagree				Strongly Agree
A Data Quality Score and star rating is relevant for industry	1	2	3	4	5
A Data Quality Score and data quality measures will help me improve data recording	1	2	3	4	5
A Data Quality Score will change how people buy rams	1	2	3	4	5
Reporting changes over time is important	1	2	3	4	5

Please tick

1. The reporting of a Data Quality Score should be publically reported: ☐ Immediately

   ☐ After a grace period

   ☐ Never

   Comments _____

2. The Data Quality Score should be reported in: ☐ RUGG reports

   ☐ Sheep Genetics website

   ☐ Other _____

3. How do you see the Data Quality Score and star ratings being used by industry?

   _____

   _____

4. Will the Data Quality Score help you think differently about data quality? (Please Circle) Yes / No

   Why / Why not?

   _____

   _____

Any other comments/ issues with reporting in RUGG and Data Quality Score reports?

_____

_____

Thank you!

10.   Appendix 2– Research output

-   2 x peer reviewed paper submissions to Association for the Advancement of Animal Breeding and Genetics (AAABG) conference. Abstracts below

CHARACTERISING THE QUANTITY AND QUALITY OF DATA USED IN MERINO SHEEP GENETIC EVALUATION SYSTEMS

S.Z.Y. Guy and D.J. Brown

Estimated Breeding Values (EBVs) published by Sheep Genetics Australia have an accuracy estimated with them. While the EBVs, their accuracy, and errors of genetic parameter estimates are all influenced by both data quantity and quality, these calculations do not explicitly take into account all aspects of data quality. To encourage increased genetic gains, Sheep Genetics provides participating breeders with data quantity and quality metrics in a 'RAMping Up Genetic gains' report. This paper demonstrates the considerable variation in these metrics for Merino flocks, and proposes additional descriptors metrics to characterise the quantity and quality of sheep genetic evaluation data. Current results show that there are opportunities to improve the completeness of pedigree and reproduction trait recording. Flocks had on average 46.6 ± 36.1% (mean ± SD) of animals with full pedigree, and 4.1 ± 6.9% of animals within each flock with reproduction trait records. The average proportion of effective progeny was 64.3 ± 19.1%. Flocks had on average 40.2 ± 37.3% of animals in contemporary groups that had variation in birth date recording. Since variation in age within contemporary groups is expected, this highlights potential issues with accurate recording of birth dates. Additional metrics describing lambing date distributions and deviations from expected were derived, and reinforce potential issues of birth date accuracy, with some flocks recording birth dates on a non-random proportion of days of the week. Feedback on the quantity and quality of their current data should help ram breeders target improvements on their recording program. However, the optimum or reasonable level of quantity and quality to maximise genetic gains is currently undefined.

MAXIMISING GENETIC GAINS WITH DATA QUANTITY AND QUALITY IN MERINO FLOCKS

S.Z.Y. Guy and D.J. Brown

Genetic gain can be maximised when selection is based on the most accurate breeding values and selection indices. To more explicitly take into account aspects pertaining to the quality of information used to estimate breeding values, metrics to characterise the quantity and quality of genetic evaluation data were previously proposed. This paper examines the relationships between these data quantity and quality metrics and genetic gains for Merino flocks. Stepwise regression analysis was used to analyse 3 genetic gains metrics: index accuracy, average index value and index trend. Index accuracy had the most number of significant predictors, with 4 quantity and 3 quality predictors explaining 85% of the observed variation. The most important metrics explaining index accuracy were level of genetic linkage for wool traits, average proportion of pedigree known in the last 3 years, and the level of wool and reproduction trait recording ($p < 0.0005$). Data characteristic metrics were also associated with average index and index trend, although to a lesser level (~24% variation explained). This study demonstrates that both data quantity and quality are associated with index accuracy and genetic gains in Merino flocks. This decomposition provides a basis for informing ram breeders on improvements in their data recording. Used in conjunction with optimum selection decisions, this will enable higher rates of genetic progress.