Livestock Genetics Investment Priorities - Industry Discussion Paper 3
October 2015

IMPROVING DATA INFRASTRUCTURE AND EXCHANGE TO INCREASE BEEF GENETIC GAIN AND INDUSTRY BENEFIT.

Purpose of this discussion paper

This is one of five discussion papers on key issues prepared to stimulate discussion of, and feedback on, the consultation draft of beef genetics research, development and extension (RD&E) investment priorities over the next 5 years.

This paper describes some of the key issues and discussion points around the current genetic evaluation infrastructure, and some of the elements required to allow the ongoing collection and sharing of data and measurements along the value chain.

The Genetics RD&E Steering Group is seeking feedback on its current perception and assessment of these key issues and on the RD&E priorities in the industry consultation draft. Feedback can be emailed to livestockgenetics@mla.com.au by 31 January 2016.

Background

One of the core recommendations arising from the industry scoping workshop in May was for ABRI to develop a single data platform or a "virtual" single database for BREEDPLAN genetic evaluation and potentially other purposes. This is driven by concerns that the way in which genetics and genomics data is currently "warehoused," protected and exchanged by multiple parties across the genetics value chain increases costs of, and limits the return on investment realised from, beef genetic improvement compared with the potential returns possible,

On a broader scale, and across many industries, data capture and communication technologies are creating new "Big Data" or "digital agriculture" opportunities, particularly where large volumes of data that are traditionally captured and housed in independent databases can be efficiently linked. This allows extraction of new information that can create potential benefits to all data owners.

For livestock breeding, improvements in animal and carcase measurement technology and gene identification will allow data to be captured on very many more relatives of animals than are currently being evaluated for genetic improvement. There is an opportunity to link the various genetic and genomic databases to the biosecurity (NLIS, LPA, NVD) and carcase feedback (Livestock Data Link, MSA) databases to capture new information, providing data sharing and privacy issues can be resolved.

MLA is currently scoping an Information Integration Program to develop this opportunity, with potential benefits to the meat industry in:

- generating real time feedback within value chains to improve efficiencies and reduce costs
- facilitating the effective exchange of regulatory and industry traceability and integrity information
- enhancing the commercial offerings of red meat processors to deliver to customer specifications

- creating better connections between producers and processors to underpin branded products
- providing the opportunity to directly link measures of carcase quality and yield into the industry genetic improvement programs
- facilitating more collaborative relationships between value chain partners and identifying new innovation opportunities
- underpinning new value chain business models and payment systems

This discussion paper discusses some options to improve data exchange used for beef genetic improvement, which can be considered as one module of the proposed Information Integration Program and an important foundation for any beef genetics value chain.

Importantly, improving data exchange should be seen as an enabling activity and not as providing a benefit per se – the industry benefit will be determined by what individual businesses and supply chain then do with the shared information, as outlined in the dot points above.

The current situation

Whilst a virtual single database or consolidated data platform linking current genetics databases has been identified as a priority, it is likely that the more important industry need that is actually being sought is the freeing up of access to data by addressing issues such as data ownership, and the protocols and incentives for sharing data.

Currently, beef seedstock breeders either retain ownership of their own herd/animal data on-farm, or they assign ownership to a breed society as part of their membership, depending on which breed society they along to. The derived pedigree and Group BREEDPLAN EBV data that results from the amalgamation of data from different herds is also owned by the relevant breed society, and is regarded as commercially-valuable information that underpins their business models. AGBU and other research organisations are able to access that data for research purposes, but only with permission from the breed societies.

Similarly feedlot and meat company managers generally believe that animal and carcase data captured in their business operations has significant commercial value which gives them a competitive edge over their suppliers and competitors.

Data ownership, and the general reluctance to share (or trade) data is at the heart of concerns that reliance on business monopolies that have been built on genetic data acquisition, but are driven by differing interests, are now limiting the genetic gains possible across the industry as a whole.

In particular, the issue of data ownership and transfer pricing needs to be addressed. Sharing data for the public good is a key strength of the current system but the financial rewards for individual data flows through the value chain are currently too low or hard to recognise, which is subsequently reflected in increased prices for data services.

ABRI currently manages pedigree/performance databases for 29 Australian beef cattle clients, and has an exclusive licence to offer a range of within-herd BREEDPLAN analyses, as well as across-herd Group BREEDPLAN analyses for 23 breed society clients, mostly run on a monthly basis.

Breed society databases holds information to underpin genetic analyses, as well as providing value adding service to its members. While the databases at ABRI are all run on the similar platform, each breed association has developed some degree of customisation. Breed associations own the data on their database and have developed these databases at their own expense.

ABRI has argued that there is no logical reason to put all of these breed society databases into a single physical database, or to "decouple" registry functions and the collection of performance data on these databases. The pedigree information on the databases undergoes significant checks and is usually subject to random DNA parent verification. The integrity of these pedigrees adds to the integrity of the EBVs produced from each BREEDPLAN analysis.

If there were to be any changes to the current beef genetics data value chain, then those changes should aim to deliver:

- Data collection processes that ensure once-only, accurate data entry
- Either a virtual network to enable data transfer and quality assurance between systems, or (eventually) a single physical database (if logistically and financially viable)
- Increased data flow and exchange between stakeholders to improve genetic evaluation and develop better information services to producers supported by financial and non-financial incentives, governance and IP arrangements, digital rights management (including ownership and licensing) and data access agreements.
- A financially-sustainable system that provides confidence to participating businesses and organisations to collectively create greater industry value through increased data access.

Key differences in genetic data collation and evaluation service models between beef and sheep.

Accurate prediction of an animal's genetic merit requires both detailed measurement of key traits that determine commercial value on as many relatives as possible and accurate recording of the specific familial relationships (pedigree) between the recorded animals. The more measurements that are available describing how related animals have performed in different environments, the more accurate a breeding value estimate of an individual animal becomes when combined with their own performance in a single environment.

Different models have evolved in the beef and sheep industries for the performance-based genetic evaluation and improvement of breeding stock, and these differences are often raised when considering where improvements can be made in each model. Because of this, the key differences between each model have been summarised in the next section.

Beef breed societies traditionally held the pedigree database and provided registration services within each breed to provide a degree of certainty for stud animals over identity, pedigree and breed "standards" (based primarily on appearance and soundness traits). The national beef genetic evaluation system BREEDPLAN has been developed since the 1980s through a "partnership" between leading performance breeders collecting performance measurements and pedigrees, MLA funding the R&D including the continual enhancement and maintenance of the analytical software engine by AGBU, and ABRI as the licensee providing commercial services to the breed societies, who act as custodians and aggregators of the performance and pedigree data

from individual breeders. The ability to use BLUP analysis to mathematically combine measurements from thousands of related individuals across generations and hundreds of stud herds that are genetically linked by use of related bulls or semen, all mapped through pedigree records underpins the great power of modern day genetic prediction.

A new role for breed societies as data aggregators has emerged with their investment in reference herds that use pooled members' funds to progeny test "elite" young sires and measure complex traits that are too expensive for individual breeders to measure in their individual studs. DNA testing in these herds allows the calibration of genomic tests that allow the prediction of phenotypic performance from a DNA sample, provided that the animals in the test herd are reasonably closely related to those in the reference herd.

Sheep breed societies traditionally held a similar role providing pedigree authentication and registration services but on a much smaller scale, as the much lower value of each breeding animal means that the cost of recording, data storage and analysis (and now genomic tests) is relatively higher than in beef. Thus when leading breeders moved into performance measurement and MLA-funded BLUP analysis was developed by AGBU, the established societies did not have the resources to manage expanding databases of pedigree and performance data, and agreed that this should be managed collectively through Sheep Genetics.

The major differences that have evolved in the structures for managing genetic evaluation data in both species are summarised below, and centre around:

- beef breed societies have a monopoly position – individual breeders can only have access to across herd data for comparative analysis and to obtain breed-specific EBVs if they join their breed society, whereas sheep breeders do not need to be members of breed societies to access across flock genetic evaluations on their seedstock through Sheep Genetics;
- The major sheep genetic evaluations are multi-breed, allowing sires to be compared on the same base across breeds within specific breeding objectives (terminal lamb production, maternal meat production, wool production).
- beef breed societies have multiple drivers within their charter, including brand differentiation and market positioning, and conflicting motivations between members within societies, whereas the Sheep Genetics service is driven only by increasing the rate of genetic improvement.

1. Data Collection

DIFFERENCE BETWEEN BEEF AND SHEEP SYSTEMS	IMPLICATIONS/CONSEQUENCES
The higher unit value for beef means that measurements, storage and analysis costs are relatively more cost effective than sheep due to relative value of the animal	• Beef breed societies have traditionally been financially stronger and more able to charge higher fees per animal recorded. • Beef has commercialised genomic services (pedigree and genotypes) and can potentially capture greater return per dollar invested in genomic tests (ie same cost per animal tested, regardless of value
Sheep have a larger effective population size per business, shorter generation intervals and higher reproduction rates than cattle	• Parentage can be more difficult to collect in sheep due to greater fecundity • Genetic selection principles can be applied more meaningfully in sheep flocks due to larger average population sizes per business • Birth weight is harder to collect in beef cattle due to animal size • Pedigree is harder to collect in northern herds due to infrastructure in pastoral environments
Beef breeders use more AI and much more imported semen.	• Reliance on overseas bulls that are marketed by semen companies means many breeders rely less on "in-house" genetic improvement, similar to dairy. • High level of AI in some beef breeds means fewer hard to collect fertility records from natural mating
Beef cattle supply chains have greater investment in measuring phenotypic performance beyond the seedstock breeding enterprise	• Beef has a much greater opportunity to collect individual animal data from commercial crossbreeding herds, feedlotters and abattoirs to inform sire EBVs through NLIS
In both species on-farm easy to measure traits are recorded by individual breeders or commercial service providers (ie scanned traits) at the breeder's cost, whereas HTM traits are collected in reference populations under different cost arrangements	• The sheep reference population is nationally coordinated, multi-breed and funded 100% by levies (via MLA). Genomic testing is run on a cost-recovery basis by the Sheep CRC • Beef reference populations are largely breed-specific and funded 50% by respective breed societies and 50% by MLA Donor Company funding (via MLA). • Beef genomic testing is coordinated via breed societies on a cost-recovery or cost-plus basis, and accessed through competing commercial services

2. Data Collation and storage

DIFFERENCE BETWEEN BEEF AND SHEEP SYSTEMS	IMPLICATIONS/CONSEQUENCES
Beef breed societies have a gate-keeper role on across- herd pedigree and performance data, and control access to breed-specific databases held by ABRI. **Sheep breeders can choose to submit data directly to Sheep Genetics or pay for a data service provider to do this for them. Sheep Genetics provides a breed-agnostic gatekeeper role, ensuring quality control of data submitted and providing feedback and results.**	• Beef breed societies have economies of scale in providing data services and managing data quality, especially around pedigree • Breed societies can provide additional services outside of data management for genetic evaluation • In some breed societies, members who do not performance-record cross-subsidise those who do. In others, those who submit data pay the additional cost of performance recording and are the only breeders to receive EBVs on their animals • Any sheep breeder who complies with data submission requirements can buy breeding value estimations, regardless of breed society membership. • In many cases independent pedigree verification services are of limited value in sheep. • Most sheep breeders choose to manage and submit their own data, and arguably take higher individual responsibility around data ownership and integrity, but at an upfront cost of greater education to achieve good data management • The sheep breeding evaluation model creates competition around data management services and commercial software services, but conversely, data services providers are fragmented and have limited economies of scale • Subsequently, sheep has a high level of engagement with commercial software services with greater reliance on software for auditing data quality rather than central system • The sheep system provides direct feedback on data quality to individuals and is simpler to make large scale uniform changes to data services, which in beef must be approved by each breed society

3. Data analysis, reporting and delivery

DIFFERENCE BETWEEN BEEF AND SHEEP SYSTEMS	IMPLICATIONS/CONSEQUENCES
For beef, breed-specific genetic evaluations are designed and purchased by breed societies from ABRI as sole licensee under fee-for-service contracts to breed societies, some individual breeders or companies (within-herd analyses) and international customers. **For sheep, evaluations are based around industry needs and breeding objectives (terminal, maternal, dual purpose) and are multi-breed - all breeds are compared on the same base within each breeding objective. ASBVs are directly comparable across breeds.**	• Beef evaluations are highly customised for genetic parameters, adjustment factors and selection indexes, which are regularly updated for some breeds, increasing complexity and maintenance costs for AGBU and ABRI – resources needed for development and refinement are fragmented. • Only genetic analyses of interest to the breed society or client are conducted, and only on data that is authorised for that specific analysis. • The majority of Sheep ASBVs are regularly re-estimated across three analytical runs, which are easily upgraded.
For both beef and sheep, a single organisation (AGBU) creates all improvements to the analytical software and ensures the integrity of the software but: • **In sheep, the same organisation (AGBU) provides the analytical service which is highly automated under contract to Sheep Genetics** • **In beef, ABRI is sole licensee to provide the analytical service.**	• For ABRI, the breed societies are their major clients, so there is less focus on services to end users, and less interest in spending time and resources on new features or enhancements that are harder to charge for. • There can be substantial delays between release of new features by AGBU and commercial release by ABRI, or in addressing resultant problems, due to resource constraints and different priorities in each organisation. • Development costs are largely breed-specific and paid for by clients resulting in higher service costs • AGBU has automatic right to use all sheep data held in database for research purposes but requires breed society permission to use beef data.
In beef, breeders pay their breed societies to organise breed-specific Group BREEDPLAN analyses, which are contracted to ABRI, the majority of extension costs are shared with MLA (via MDC funding), the majority of the development costs associated with delivery, but none of the research costs, unless there are highly breed-specific outcomes required. **In sheep, breeders pay directly for the routine genetic evaluations they require, the majority of the extension costs (included in data submission costs), but none of the associated R&D costs of delivery**	• Routine beef evaluations are delivered via fully commercialised monopoly services with limited additional funding, but with increased cost and complexity, and a longer pathway to delivery through multiple organisations (and multiple negotiations for changes) resulting in slower delivery of R&D outcomes from AGBU to end users. • The sheep evaluation service is owned by the industry, with the breeder end user as client, so there is greater transparency and accountability for customer service and timely implementation of industry R&D outcomes (via an end user advisory committee) and for issue resolution (via an independent technical committee) • Conversely, because the business scale is not large enough to be fully commercialised, there is no commercial incentive to do overseas evaluations (although NZ and USA breeders are clients), the service is dependent on RDC funding, and the service is also a monopoly service with limited breeder ownership.

Potential enhancements considered by the Genetics Steering Group

There are a number of data repository and "big data" management concepts under active consideration for the meat and livestock industries. The Genetics Steering Group has been made aware of, but makes no judgements on, several of these alternative models – each requires more detailed development through discussion with contributors and users to determine relative strengths, weaknesses, practicality and sustainability.

1. Global cross-referencing to enabling a virtual single database

Each time ABRI runs a BREEDPLAN analysis that includes data from more than one database, eg a trans-Tasman analysis, a "transitory" virtual single database is established, containing only the subset of information required for genetic analysis, rather than all of the information that is contained on each breed association database. This maintains the ownership status of the data.

Creation of the single database requires cross referencing of animals which are common across the two or more databases. The advantage of this is that the virtual database has the most up to date information available across the databases involved, while allowing the autonomous databases to collect the information they are interested in. Each participating database gets the genetic outputs (EBVs, indexes, etc) back on their own animals.

ABRI has proposed the creation of virtual single databases would be enhanced by the creation of a "global cross referencing database" that was available to database providers to cross-reference animals in their database to other databases. The cross-references would then be stored in the central repository.

This global cross referencing database would be used as a very efficient way of creating "virtual" single databases each time they were required for genetic analysis, allowing for an unlimited variety of identification structures, which is extremely important when dealing with established databases with historic animals or where strict regulations describe the format of the identification – particularly overseas where National Identification may be legislated. The global cross reference database would be cloud based with web enabled access strategies.

The very strong advantage of a global cross-reference system is that it can be implemented without the need to get unanimous and universal agreement. It simply needs to be set up and made available.

Databases like NLIS and MSA would not use the global cross-referencing system. Rather, the NLIS identification needs to be stored on the other databases if the link was to be established. Many Australian databases already record individual animal NLIS identifications as an option. Similarly, databases in other countries record their own "national idents" that may also link them to their own industry data.

Whilst addressing a significant issue for simplifying analyses that require retrieval of data from multiple databases within ABRI's business, this approach does not address the concerns pertaining to complexities and barriers to data access and to business models that are not necessarily aligned to maximising genetic gain

2. National Data Repository

The Australian dairy industry has recently moved to reposition dairy data exchange into a precompetitive setting that is seen as essential to support research, improve genetic evaluation and underpin information product innovation across that industry[1]. The concept includes herd management and product quality data, so is more akin in scope to the Information Integration Program promoted by MLA.

The main challenges identified with the current industry infrastructure and data exchange identified included:

1. Lack of industry leadership to address data issues and realise productivity gains
2. Multiple animal ID processes and numbers with calls for implementing a unique recognised animal ID system
3. Gaps in data; value adding constrained
4. Fragmented systems, difficulties in data transfer and sharing
5. Less than optimal data collection processes and adoption of new technologies
6. Reduced data flows through the data value chains
7. Improvements in reliability of genetic evaluation
8. Data not valued by all stakeholders
9. Incentives for data collection not aligned with benefits of data use.

Their vision to "create a centralised industry-owned repository where quality-assured data from all sources are available for industry-wide use" has already been achieved, at least for sheep industry genetics data, by the Sheep Genetics model within the meat industry.

3. Pre-competitive data analysis

An alternative model has been proposed that introduces three significant elements in a new infrastructure and data exchange model to reduce complexity and cost for the beef genetics sector:

1. A "pre-competitive" national single database (NSD) for storing phenotypes and genetic analysis.
    - Data in the NSD is collated for industry use.
    - Audit software scrutinises data received and imports only quality information.
    - The NSD applies biological and IP protected genetic parameters.
    - The NSD produces EBVs based on agreed industry-developed genetic parameters and provides these to service licenced providers.
2. Licensed service providers can access the NSD and provide the following
    - Collect and submit client breeders phenotypes
    - Report EBV produced by NSD to their clients

---

[1] National Herd Improvement Association "Dairy Industry Data Working Group Report for NHIA Dairy Industry Data Project Final Report" July 2010

- Provide specific solutions for client breeders from data provided from the NSD that has had agreed biological and IP protected genetic parameters applied.
3. Phenotypes are valued so that the industry can financially stimulate trait collection by:
   - Creating a commercial value for each trait.
   - Crediting the service providers for the value of each trait that passes audit.
   - Debiting the service providers for each trait's genotypic result.
   - Creating a financial incentive to service providers to train and supply value-added services to breeders.

Note: licence conditions will be crucial in determining data access and exchange rules – and therefore the ways in which behaviours and business models are changed.

4. Drawing on private sector expertise in managing supply chain data exchange

A number of companies have successful business models based on developing and maintaining customised data management business solutions for very large, multinational supply chains in agriculture. These companies have considerable expertise in addressing all of the current concerns raised in discussion around exchanging on-farm, feedlot and carcase performance and genomics data including, ownership and IP, valuation, standards, encryption, limited access, etc[2].

Given the crucial importance of this issue, the Genetics Steering Group recommends that the coinvestors in the genetics delivery system draw on this commercial experience in developing systems and processes to protect and share data between businesses in any future modification of the genetics value chain.

5. The Livestock Information Platform concept (under consideration by CSIRO and ABRI).

Universal availability of high speed and reliable broadband is about to occur throughout rural areas with the commissioning of two high-capacity NBN satellites to be launched from October 2015. This development combined with virtually unlimited and scalable "Cloud" computing, data storage and automated capture of on farm data will, for the first time, enable a **birth to butcher** story of livestock to be recorded in detail and analysed to extract maximum value.

Wherever an animal is, whenever it travels regardless of whoever owns or manages it, its story can be traced with the help of LIP. Everyone along the supply chain, whether producer or processor, can work together by adding value to their business and their industry with the shared information that LIP could provide.

Any producer, processor or industry body with identifiable animal data may deposit the data into LIP. There is no restriction upon what the data may represent. Data depositors may use existing data descriptions, or create new ones. Data deposited will remain the property of their owners. Owners may choose to keep their data absolutely private, or they may wish to make it totally public. Variations between these two will include

---

[2] An example is F4F Agriculture (www.f4f.com) which provides both expertise and a technology solutions framework that is integrates all parts of the agricultural supply-chain and providing custom data-management business solutions.

publicly available aggregated data only, or individual animal data, but with anonymous identifiers. Owners may also choose to restrict access to their data to trusted individuals or organisations.

Access to the data by Applications (Apps) will be possible via an openly published Applications Programming Interface (API). Apps, like individuals, will need permission from data owners to access the data. LIP will be jointly developed by computer and livestock scientists from ABRI and CSIRO with proven track record in developing livestock and cloud systems. ABRI and CSIRO will invest in the project.

The system would be developed using an integrated set of Microsoft tools and cloud computing infrastructure. These platforms are easily scalable, globally available (including Sydney and Melbourne) and extremely reliable. A range of tools are now available that specifically and efficiently cater for the varied nature of the 'big data' that would be being brought together under the LIP umbrella.

LIP would also be designed to be ready for the emerging 'Internet of Things' (IoT) technology. Apart from the commonplace PC's, tablets and smart phones, interfaces can be developed for specialist devices. These might include hand-held instruments, animal mounted sensors or automatic remote sensors whose data is sent via wireless, mobile or satellite. Previously incompatible applications would speak the same language and become compatible within the LIP environment. This cross data-source compatibility would allow development and implementation of Apps that interrogate, analyse, interpret and control components of whole farm systems. Some Apps may be custom built but many would be suitable for multiple uses.

Discussion points

1. *To what extent will the global cross-referencing database proposed by ABRI address the real needs underlying the call for a single virtual database?*
2. *What value will it have other than for facilitating BREEDPLAN runs?*
3. *Whilst focussed initially on linking genetic information databases, does it have wider benefits for accessing other information in future?*
4. *Is there a mechanism by which input data from individual breeders can be valued and traded for output data from combined analyses provided through the breed society to encourage measurement of "rare" or hard-to-measure traits, particularly as DNA tests make performance measurement less attractive?*