DATA STANDARDS FOR THE AUSTRALIAN SHEEP INDUSTRY

T.R. DYALL[AC], D.A. ECCLES[BC], D.J. MIRON[AC] and R.D. NETHERY[AC]

[A] CSIRO Livestock Industries, F D McMaster Laboratory, Armidale, NSW 2350
[B] NSW Agriculture, Forest Rd, Orange, NSW 2800
[C] Australian Sheep Industry CRC Project 2.2.3, Locked Bag 1, Armidale, NSW 2350

SUMMARY

The Australian sheep industry is evolving into a high technology industry requiring large amounts of data collection. Sheep producers increasingly rely on electronic recording of information to improve productivity. Currently, there are many disparate data management applications involving different terminologies and data architecture, and no standard data transmission protocols. There are a bewildering variety of software and data formats from which to choose, but there are no standard data formats to establish a common ground across the sheep industry. This has led to confusion, but more importantly, is a key barrier to the uptake of decision support software within the sheep industry. The Sheep Industry Data Standards Project provides a standard set of data terms and protocols so that data management and transfer can occur without confusion across the sheep industry.

*Keyword*s: data, dictionary, Extensible Markup Language (XML), messaging, standards, sheep

INTRODUCTION

Electronic recording of information is becoming more prevalent among sheep producers, allowing them to use the data to increase productivity and profitability. Electronic data recording requires Data Management Software (DMS). There are now a number of vendors within Australia offering DMS ranging from the simple to the complex, either as stand alone or Internet-based applications. Not all producers have opted for specialised DMS. Many prefer to simply store their sheep data electronically and send it via email to a consultancy for analysis. Over time, each DMS vendor has developed their own Database Architecture and Data Dictionary for their clients. This independent development by DMS vendors has led to confusion among producers. For example, many DMS vendors use *micron* to represent fleece diameter, while other vendors use *fleece diameter*. Similarly, many wool measurements have been abbreviated, but the abbreviations are not consistent across the industry. For example, Coefficient of Variation of Fibre Diameter may be represented as CV, CVD or FDCV. Data field names have been developed *ad hoc*, with minimal adherence to conventions or common rules used in creating databases. Field names may be a full name or an abbreviation; contain spaces or characters (e.g. slashes or commas); appear as capitals or lower case, or a mix of the 2. Hence, DMS vendors present many data field names differently.

The lack of standardisation within the sheep industry costs software vendors, service providers and sheep producers time and money. Moreover, it is reasonable to suggest that this lack of standardisation discourages producers from considering an electronic sheep data management system. At the e-sheep® Forum held in May 2003, by the Australian Sheep Industry CRC, producers and vendors alike called for a standard for the storage and transmission of on-farm sheep data. This standard was to include a data architecture, data dictionary and an Extensible Markup Language (XML) messaging specification. In parallel, the Australian Wool Innovation (AWI) had also identified a need for a similar standard (McDonald and Hansford 2003). As a result, the AWI, in conjunction with the Australian Sheep Industry CRC, funded this data standards project to generate such a standard.

MATERIALS AND METHODS

When a producer has data analysed commercially, the service provider has to map the producer's on-farm data in a readable format. Following analysis, the results are then typically returned in a format unreadable to the client's DMS, requiring the vendor to produce filtering software. If the client uses a spreadsheet for data management, spelling mistakes and data errors can occur. This requires direct communication between the provider and client to resolve the problem. Standards have not been developed for file transfer so there is a wide range of file formats in use by DMS vendors. For clients with limited computer skills, this can cause difficulty when importing files from, or exporting to, a database, especially if a number of DMS vendors (using different file formats) are involved, or if the

producer changes vendors. Files may be transmitted using many data formats including: CSV, ASCII comma delimited, ASCII space delimited, text, Microsoft® Excel or XML.

Extensible Markup Language
Extensible Markup Language is a meta language (see http://www.w3c.org/) for the description of structured data and documents. It is designed to describe data, and focus on what the data are, for example, a column in a table or a fleece on a sheep. It is not HTML, which is focussed on how to display data. Originally designed for large-scale electronic publishing, it is now being utilised for the exchange of electronic data. It enables a common representation of data, which makes it ideal for data exchange between disparate systems. Given the enormous variety of synonyms in use within the sheep industry, XML provides a flexible, reliable solution for exchanging data between existing DMS. As an example, a table called 'Paddock' in 1 system may have the columns and row data as follows:

pdkPrimaryId	ppyPrimaryId	Area	Name	Pasture_Description
1	34	300	Tank	Phalaris

Another system (US in origin) may call the table 'Field':

PaddockKey	PropertyKey	Size	Name	Pasture
1	34	300	Tank	Phalaris

Importing the first set of data into the second just needs a mapping of the XML output to say that pdkPrimaryId is now PaddockKey and so on.

Two other major factors contribute to the confusion and lack of standardisation within the industry. Firstly, On Farm Fibre Measurement (OFFM) is serviced by 2 different measuring systems - the Laserscan and Fleecescan technology developed by CSIRO and AWTA, and the Optical Fibre Diameter Analysis (OFDA) technology developed by the Interactive Wool Group company. These technologies vary in the way they calculate, report and name a range of wool fleece measurements. For example, using OFDA, coarse edge micron is defined as the difference in microns between the mean diameter and the broadest 5% of fibres. Using Laserscan, coarse fibre content is calculated by determining the percentage of fibres in a fibre diameter distribution that are coarser than 30 microns. Secondly, sheep genetic evaluation has evolved through a number of different service providers, resulting in different naming and measurement standards for the same trait. For example, the trait CF may mean Comfort Factor measurement of the fleece or Carcass Fat measurement of the animal.

RESULTS
The Data Standards Project has produced a set of standards that will provide the industry with a common language for on-farm data storage and messaging. Three components have been created:
- A data architecture to represent the structure of a reference on-farm database implementation.
- A data dictionary to define the field names used in the data architecture. The definition of a field name includes not only the field name, but also its data type and its meaning.
- An XML messaging specification for the transmission of messages between producers and service providers. The areas identified where messaging is required are classing, testing, genetics, carcass data and animal health.

Creation of these components required the input of DMS suppliers and service providers nationwide (Figure 1). A major consideration was how the desired standard would interface with other existing standards. In order to produce an industry-accepted standard, it was crucial that industry had major input into the standard's development, and was able to reach consensus on the contents of the components.

Given the sensitivities of the commercial parties and their products, this input and consensus was achieved by ensuring that the Data Standards Project presented a neutral and unbiased position. This impartial stance was accomplished by creating a project structure consisting of a Management Committee and a Combined Working Group whose members were from independent organisations. The Management Committee was made up of 3 representatives, 2 from CSIRO and the other from AWI. Its role was to resolve any issues arising, technical or otherwise, throughout the project. Under the Management Committee, a Combined Working Group was created consisting of members of the

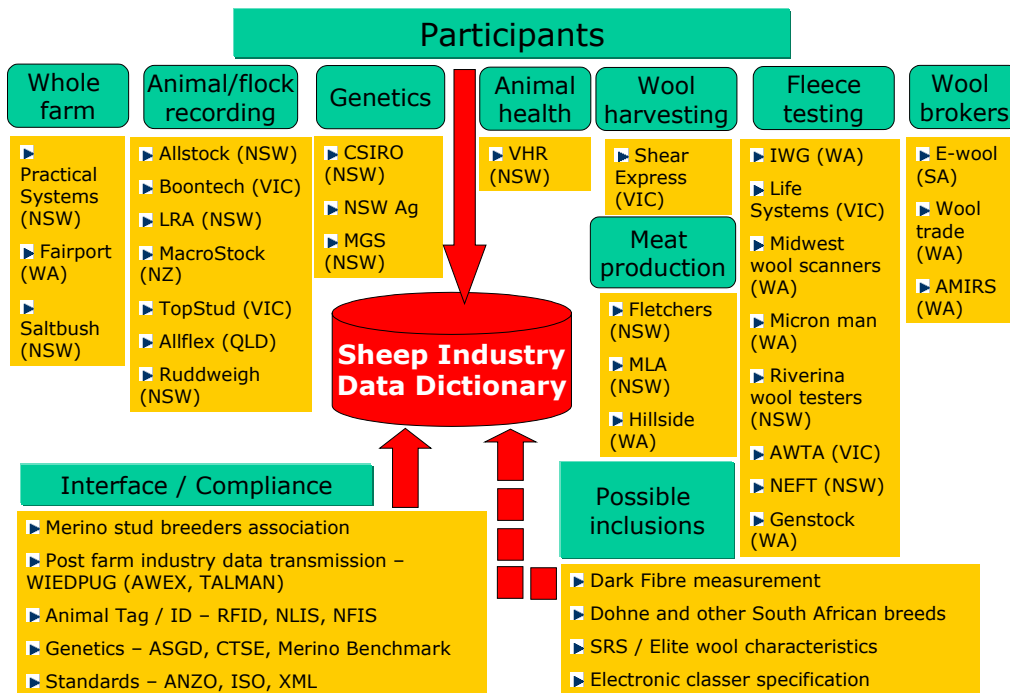Management Committee and other members from CSIRO, NSW Agriculture and Meat and Livestock Australia.

Figure 1. Participants in the sheep industry data standards project.

To give the vendors and service providers the opportunity to participate, Expressions Of Interest (EOI) were called for from throughout the industry. An affirmative response to the EOI included the supply of the organisation's current data dictionary and data architecture. Some forty parties were contacted with over 70% responding positively, supporting the initiative. Responses to the EOI demonstrated a significant disparity between vendors' implementations of data architectures and data dictionaries. As would reasonably be expected, the commercial providers insisted that their responses were kept confidential.

DISCUSSION

Key recommendations of McDonald and Hansford (2003) to address the shortcomings of current decision support systems (DSS) were:

- Definition of industry transmission standards.
- Development of a data dictionary, data architecture and a reference database for the sheep industry.
- Data dictionary, database and transmission standards to be freely available on a website.

To address these recommendations, this project extended invitations to software developers and industry bodies to participate in the development of the standards. The involvement of 70% of those invited, and the responses from them, were a clear endorsement of the immediate need for the development of data standards for the Australian sheep industry. As a key component of the project, a number of information gathering exercises were undertaken with individual software developers, and an industry forum was held at CSIRO, Armidale, in December 2003. Feedback from the participants focussed on the notion that standards needed to be implemented as soon as possible, to aid in the faster uptake of DSS systems by sheep producers, and to aid in the further development of their software systems. A key cost to the software developers was the down time spent with sheep producers in helping them transfer data files with correct information, and in the right format.

This project has successfully created a set of data standards for adoption by the sheep industry. This will reduce the cost of doing business within the industry through consistency of terminology and a common messaging protocol. The standards will also reduce confusion amongst producers, enabling them to approach DMS with confidence. The data transmission standards have been developed using XML, because it is ideally suited for web-based systems and can be adopted by disparate systems. A draft XML specification for an electronic wool classing specification sheet was developed and further

documents have been flagged for later development. Many files are currently transferred using CSV format, but the majority of software developers wanted standards developed using XML as it is more flexible and is being adopted by the Australian Wool Exchange for post-farm transmission of wool data.

Adoption of standards for the Australian sheep industry will depend on the willingness of software developers to adapt their product to incorporate the standards. Adoption will also depend on the market forces of sheep producers demanding software for on-farm data storage and transmission that is compliant with the industry standards. At present, there is no dominant player or industry group that can force their standards onto the industry. The standards will encourage new players to enter the market and this may persuade current software developers to adopt the standards. Some of the software developers involved in the project are primarily focussed on cattle producers, and are keen for standards to be put in place to aid them in quickly adapting their systems for sheep producers.

The next steps of this project are to:
- Promote the data standards within industry through an extension program.
- Set up an industry body to manage the ongoing development and implementation of the standards. This body would take over the role currently being carried out by the Combined Working Group.
- Develop XML messaging standards for common industry documentation such as an electronic wool classer specification, vendor declarations (chemical and dark fibre risk), and wool test data from small wool test houses.
- Seek International Standard Organisation and Australia and New Zealand Organisation approval.

ACKNOWLEDGMENTS

REFERENCES
MCDONALD, M. and HANSFORD, K. (2003). 'Analysis of OFFM Decision Support Systems.' (Final Report to Australian Wool Innovation Ltd.)

Email: tim.dyall@csiro.au