



Farming out data-driven decisions

Data ownership and security are topics dairy must address in order to stimulate innovation.

by the University of Wisconsin-Madison Dairy Brain's Coordinated Innovation Network (CIN)

DATA has played an integral role in dairy farmers' decision-making process for many decades. Much of this started with foundational work from land-grant universities and state extension services. Groups such as National Dairy Herd Information Association (DHIA) and USDA's Animal Improvement Programs Laboratory (AIPL) paved the way generating production and genetic metrics. More recently, AIPL was rebranded into a research-only entity, now the USDA-Animal Genomics and Improvement Laboratory (AGIL), while the Council on Dairy Cattle Breeding is now responsible for genetic evaluations.

For the good of all

A likely cause of early successes, perhaps unintentional, was that the dairy producer had full control over who he or she provided data to.

The dairy industry trend suggests that even more expressive data sets will be available in the not-too-distant future. In fact, advancements in cloud-based data systems and advanced analytics (machine learning, optimization, and visualization) suggest that there will be a wide gamut of exciting opportunities for the farmer to utilize data to positively impact on-farm management. Maintaining the producer-centric data control model should form the cornerstone of collaboration, but there will undoubtedly be important lessons to learn as the scale of data expands.

This article continues the discussion from a subgroup of the University of Wisconsin-Madison Dairy Brain's Coordinated Innovation Network (CIN). This USDA-funded effort is multifaceted, but one important goal is to stimulate a larger and more diverse industry discussion regarding data ownership and security issues for both today's and tomorrow's problems.

Being able to scale the industry's data systems first requires that the industry tran-

sition to modern methods of data exchange according to international standards. The driving organization involved in standardization in the dairy industry is the International Committee on Animal Recording or ICAR. One effort led by ICAR is standardizing the key data communication through the release of their Animal Data Exchange Standards.

Exchanging data through standard files can offer structure, error checking, and other benefits that will be necessary as the volume of data and metadata morphs — the old way of sharing data through downloading flat files has ultimately proven to be cumbersome and error-prone. The ICAR data format standard is a relatively recent contribution and is illustrative of the speed at which the industry is moving toward cloud-compatible data systems that encourage easy transmission of data through programs such as VAS Platform, Bovisync, and the various milking systems.

Who, what, and when

Data portals are being developed utilizing these data exchange standards in order to facilitate the transmission and use of data between entities. Under current circumstances a customized data use agreement is signed by both entities that details how the data is to be shared, utilized, and who can have access to the data.

This manual process might work for today's data pipelines, but it may soon be advantageous to automate this process and have data portals where a dairy producer can simply add and subtract users of their data on a website form. These portal improvements will go a long way to encourage responsible use of data, but there are other questions about data and ownership that must be discussed openly, specifically around the topics of derivative products and data persistence.

Data portals might ease the pain of granting access to data streams, but if new intellectual property (IP) is added to a raw data set, the derived works could be considered to be a new, independent product with its own associated rights and permissions. Without the appropriate language in a data use and exchange

THERE ARE MANY LAYERS OF COMPLEXITY to data ownership and data security that can be difficult to navigate as a dairy producer or allied industry professional.

agreement, the potential exists that competing interests are created between the original data generator and vendor. While this paradigm might suffice for today's data uses, the next generation of analytics demand huge data sets combining data from different sources — there is a growing recognition that there is value to data.

Chain of custody

The concept of a data portal might need to be extended to also establish a "chain of custody" for a particular set of data. A chain of custody would transparently link the owner of the data to the user of the data. Once this chain of custody is established, it will be important for data generators to understand how long their data might reside within a derivative data product.

The European Union's General Data Protection Regulation (GDPR) recently established privacy guidelines that ensure a process whereby an entity can request the "right of erasure." It is an open question whether this level of privacy protection is appropriate for the U.S. dairy industry — being able to establish the next generation of dairy data systems requires careful planning on the part of all market entities.

Sensitivity of data, meaning what risk might entail if control over data were lost, is a common concern that extends to nondairy data that individuals or businesses generate daily. It is reasonable to consider that different data streams have different levels of sensitivity.

Would financial data be more or less sensitive than herd management data? What is the justification thereof?

Taking the time to evaluate the sensitivity of different data streams on the dairy will help provide a reasonable approach to data sharing and security concerns. Security should be proportional to the sensitivity of data, meaning greater efforts would be placed on securing and protecting data of greater sensitivity and vice versa. Therefore, in order to establish appropriate measures of security, sensitivity must first be addressed. One approach to evaluating sensitivity is to compare dairy data to the sensitivity of data that we generate or control in our lives.

Moving the dairy industry forward

The Dairy Brain CIN provides not only a coordinated effort to provide resources for data collection, harmonization, and analytics, but it is also a forum to discuss important topics pertaining to data ownership and security that both dairy producers and allied industry are currently facing. As an industry, we need to consider the points raised in this article in the hope of establishing modern practices pertaining to data ownership and security and stimulating data innovation. 🐄

ARTICLES FROM THE DAIRY BRAIN TEAM

- February 10:** Make better use of data
- March 25:** Data ownership and security
- April 10:** Data collection and communication
- April 25:** Making data work on the farm
- May 10:** Creating value from data



Share your input with the Dairy Brain team

To contribute to the discussion, go to on.hoards.com/DairyBrainForum or scan the QR code with your smartphone.



The collaborators and authors of this article include Adam Christensen, Derek Cooper, Robert Fourdraine, Benjamin Griffiths, Chris Mathis, James Koltes, Allison Quick, and Matthew Utt. Project supported by USDA-NIFA-FACT grant 2019-68017-29935.