The US Dairy Brain Project: Data integration and data applications for improved farm decision-making

L. Fadul-Pacheco[1], V.E. Cabrera[1], S. R. Wangen[2], T. da Silva[1], F. Zhang[1], R.H. Fourdraine[3] and J.M. Mattison[4]

[1]University of Wisconsin-Madison, Department of Animal and Dairy Sciences, 1675 Observatory Dr., 53706, Madison, Wisconsin, USA
[2]American Family Insurance Data Science Institute and Wisconsin Institute for Discovery, University of Wisconsin–Madison, 330 N Orchard St., 53715, Madison, Wisconsin, USA
[3]Dairy Record Management Systems, 313 Chapanoke Road, Suite 100, 27603, Raleigh, North Carolina, USA
[4]Dairy Herd Improvement Association, PO Box 930399, 53593, Verona, USA
Corresponding Author: lfpacheco@wisc.edu

Abstract

Data pervades the dairy farming industry. However, specific data streams are most often ad-hoc and poorly linked to each other and to decision making processes. Dairy farms have embraced large and diverse technological innovations such as sensors and robotic systems, and can now stream vast amounts of data, but they have not been able to integrate all these data effectively to improve whole-farm decision making. Consequently, the benefits of the new smart dairy are not being fully realized. In order to address this, it is imperative to develop a system that can collect, integrate, manage, and analyze on- and off-farm data in real-time for practical and relevant analyses that can be used to improve on-farm decision making. Hence, we are developing a real-time, data-integrated, data-driven, continuous decision-making engine: The Dairy Brain by applying Precision Farming, Big Data analytics, and the Internet of Things. This is a trans-disciplinary research and extension project that engages multi-disciplinary scientists, dairy farmers, and industry professionals. We are using the state-of-the-art database management system from the University of Wisconsin-Madison Center for High Throughput Computing to develop our Agricultural Data Hub that connects and analyzes cow and herd data on a permanent basis. This involves cleaning and normalizing the data as well as allowing data retrieval on demand. We have a four-part strategy: (1) Create a Coordinated Innovation Network (CIN) to shape data service development; (2) Create a prototype Agricultural Data Hub (AgDH) to gather/disseminate multiple data streams relevant to dairy operations; (3) Build the Dairy Brain – a suite of analytical modules that leverages the AgDH to provide insight to the management of dairy operations and serve as an exemplar of an ecosystem of connected services; and (4) Design and execute an innovative Extension program. We illustrate our Dairy Brain concept with a practical application that predicts CM onset. The application uses machine learning algorithms to identify cows at higher risk of contracting CM seven milkings before the onset. The application integrates data from management software and data from the milking parlor. Our preliminary results indicate that our predictions are 72% accurate. Integration of more data streams and incorporating larger historical datasets will improve accuracy even further. We demonstrate that it is possible to develop integrated continuous decision-support tools. Tomorrow's dairy industry will be built on the effective capture and integration of more data streams, not fewer. This is a critical moment to develop the structures that can

move the industry towards modernized data exchange. This is an ongoing innovative project that is anticipated to transform how dairy farms operate.

Keywords: Data integration, decision support tools, dairy brain.

Introduction

Dairy farms generate large amounts of data. With the adoption of new technologies these amounts of data have increased. These data are an important tool to help improve farm decision making in animal welfare, performance, and long-term sustainability of the farms (Lovarelli *et al.*, 2020). Nevertheless, using data in efficient way is a complex task as it is highly heterogenous, different data sources are normally independent from each other (Wolfert *et al.*, 2017) and the analysis of big data is a challenging task (Morota *et al.*, 2018).

As a result, daily management is a difficult task since data integration of different data sources (i.e., production, feeding, health, etc.) is lacking (Koltes *et al.*, 2019; Cockburn, 2020; Cabrera and Fadul-Pacheco, 2021). It has been recognized that collection, aggregation and analysis of data will transform the way decision making is done, making farms realize large gains in productivity, efficiency and profitability (Bronson and Knezevic, 2016; Newton *et al.*, 2020) and providing farmers a better understanding of the past, present and future of the farm (Lioutas *et al.*, 2019).

Data integration is crucial to improve data quality and algorithm performance (Hogeveen, *et al.*, 2010; Menéndez González *et al.*, 2010) and it has been identified to be an important component to the decision-making process on dairy farms (Eastwood *et al.*, 2017; Dairy Brain, 2020a). In addition, automatized data integration is recognized as a tool to give holistic advice on management practices (Gengler, 2019).

Knowing the potential benefits of data integration, it is imperative to develop a system that can collect, integrate, manage, and analyze on- and off-farm data in real-time for practical and relevant analyses that can be used to improve on-farm decision making. Hence, we are developing a real-time, data-integrated, data-driven, continuous decision-making engine: The Dairy Brain by applying Precision Farming, Big Data analytics, and the Internet of Things. This is a trans-disciplinary research and extension project that engages multi-disciplinary scientists, dairy farmers, and industry professionals.

The Dairy Brain

The Dairy Brain project as a continuous decision-making engine as described on Figure 1, first data is collected at the farm, then farm data from multiple sources is transferred to a central location, where data is transformed and homogenized to then apply analytics to it and finally access to the developed analytics farmer via web interface. It consist of a four-part strategy: 1) Create a Coordinated Innovation Network (CIN) to shape data service development; 2) Create a prototype Agricultural Data Hub (*AgDH*) to gather/disseminate multiple data streams relevant to dairy operations; 3) Build the *Dairy Brain* – a suite of analytical modules that leverages the aggregation service and available data to provide insight to the management of dairy operations and serve as an exemplar of an ecosystem of connected services and 4) DataMoney: an innovative Extension program.

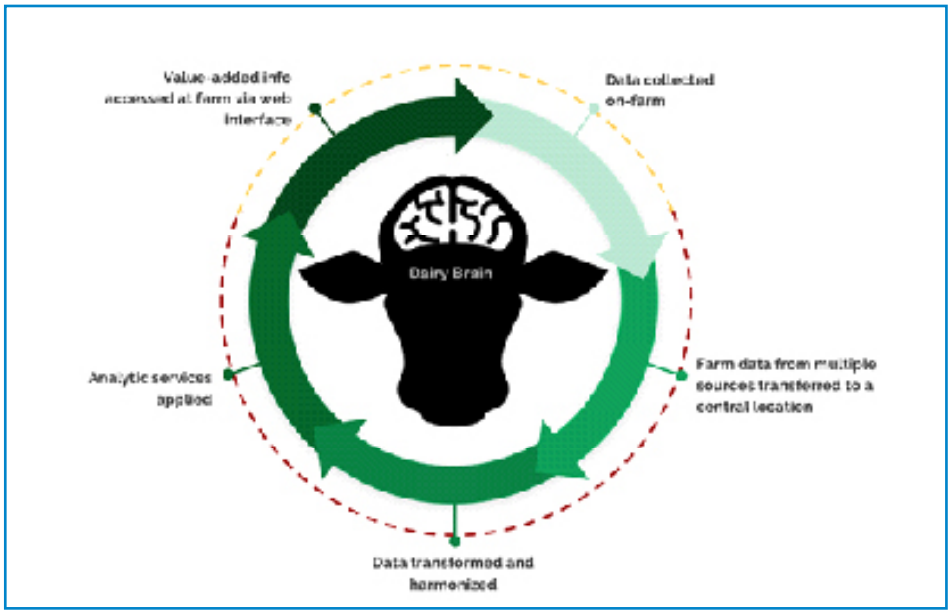

Figure 1. Dairy Brain project phases.

Create a Coordinated Innovation Network (CIN)

A Coordinated Innovation Network (CIN) is, in general, "a large community of stakeholders that addresses bottlenecks in critical areas by bringing together experts from different disciplines and domains to identify innovative and synergistic solutions" (USDA, 2021). In the context of the Dairy Brain project, the *bottlenecks* make reference to the data management challenges in the dairy industry. Therefore, the CIN of the dairy brain project is a network of stakeholders from the dairy industry (i.e., farmers, industry, researchers and extension professionals) shaping the structure of the Dairy Brain project itself and serving as a basis for broader implementation of data services and standards.

The main roles of the CIN are to raise awareness, facilitate the exchange of opinions and generate discussion and crate guidelines about data management in the dairy industry. To start the discussion process, the CIN published five opinion articles that covered some data management challenges that were identified by the CIN members and the Dairy Brain team.

The first was an introductory article to the Dairy Brain project and the roles of the CIN (Dairy Brain, 2020a). This was followed by a discussion about data security and data privacy (Dairy Brain, 2020b). The next article was about data collection and data standardization (Dairy Brain, 2020c). Then, another article was the adoption of decision support tools (Dairy Brain, 2020d), and the last opinion article was related to the value added of data to help improve management practices at the farm (Dairy Brain, 2020e). These opinion articles are the foundation for more technical and scientific articles: the CIN design documents, which are under development. To learn more about the perception of a number of key topics related to data challenges in the dairy industry, we are currently conducting a survey at DairyBrainSurvey (*https://uwmadison.co1.qualtrics.com/jfe/form/SV_0HDBzawvvRygQVE*).

Create an agricultural data hub (AgDH)

Farms have different data sources that are normally not connected to each other, to address this issue, the main objective of the agricultural data hub (AgDH) is to collect, integrate and homogenize the dairy farm data so that all these data can be used by the DairyBrain to develop decision support tools. Data integration through the AgDH consist in five main steps: 1) Assessing; 2) Decoding; 3) Cleaning; 4) Homogenization and 5) Integration. The AgDH is built in a modular fashion under the consideration that these five steps will be automatized with minimal or no human supervision. Once the data is fully processed, these integrated data will be accessible through a set of application programming interfaces or API's endpoints hosted by the AgDH service. It is important to mention that these endpoints will be secure so that only individuals authorized by the farm will have access to them. Also, the API's will make the data available to the analytical module offered though the DairyBrain portion of our project.

Build the DairyBrain

The actual Dairy Brain portion of the project is a data-driven engine of decision making to advance analytics and dairy farm sustainability. We categorize our models as descriptive, predictive and prescriptive. The descriptive models are normally for short-term decision making, and they are mainly visualizations (i.e., dashboards). They might also include some simple calculations, as for example, feed efficiency. Even though feed efficiency is based on simple algorithms, it requires data from different data sources which means it needs data integration.

Predictive models include performance projections to the future. One example of this type of models is selection of genetic traits to reduce the incidence of clinical mastitis (Fadul-Pacheco *et al.*, 2021). And, finally, the most advanced, the prescriptive models are those that provide suggestions, mostly from optimization, of the best course of action. Among these models we can find the continuous nutritional accuracy (Barrientos-Blanco *et al.*, 2020) that can help provide accurate diets to cows as an effective strategy to control cost, increase revenue, enhance feed efficiency and reduce environmental impact. It is important to keep in mind that all these models need to be adjusted with continuous data. Data integration from multiple data sources though the AgDH and the application of advance analytics is an example of added value to the raw and disparate data.

DataMoney: An innovative extension program

The objective of the extension program of the Dairy Brain project is to increase farmers, advisors, and county educators' awareness on data use and management at the farm level. More precisely, demonstrate the benefit of data integration and the appropriate use of decision support tools. We have developed a program called "DataMoney." The program consists in working individually with farm teams in particular farms. The first step is to perform an assessment of the data usage at the farm and, according to the farm priorities, needs assessment, and farm data availability, work collaborative developing farm-specific decision support tools. This process will spark farmer and farm workers interest in the use and application of farm data for better decision making.

List of references

Barrientos-Blanco, J.A., H. White, R.D. Shaver, and V.E. Cabrera. 2020. Improving nutritional accuracy and economics through a multiple ration-grouping strategy. Journal of Dairy Science 103:3774–3785. doi: 10.3168/jds.2019-17608.

Bronson, K., and I. Knezevic. 2016. Big Data in food and agriculture:. Big Data and Society. doi:10.1177/2053951716648174.

Cabrera, V.E., and L. Fadul-Pacheco. 2021. Future of dairy farming from the Dairy Brain perspective: Data integration, analytics, and applications. International Dairy Journal 105069. doi:10.1016/j.idairyj.2021.105069.

Cockburn, M. 2020. Review: Application and Prospective Discussion of Machine Learning for the Management of Dairy Farms. Animals 10:1690. doi:10.3390/ani10091690.

Dairy Brain. 2020a. Help Us Help You Make Better Use of Dairy Data. Accessed September 24, 2020. *https://hoards.com/article-27981-help-us-help-you-make-better-use-of-dairy-data.html*.

Dairy Brain. 2020b. Farming out Data-Driven Decisions. Accessed September 24, 2020. *https://hoards.com/article-27982-farming-out-data-driven-decisions.html*.

Dairy Brain. 2020c. Data: Think Big, but Start Small. Accessed September 24, 2020. *https://hoards.com/article-27983-data-think-big-but-start-small.html*.

Dairy Brain. 2020d. Making Data Work on the Farm. Accessed September 24, 2020. *https://hoards.com/article-27984-making-data-work-on—the-farm.html*.

Dairy Brain. 2020e. Creating Value from Data. Accessed September 24, 2020. *https://hoards.com/article-27985-creating-value-from-data.html*.

Eastwood, C.R., B.T. Dela Rue, and D.I. Gray. 2017. Using a 'network of practice' approach to match grazing decision-support system design with farmer practice. Anim. Prod. Sci. 57:1536. doi:10.1071/AN16465.

Fadul-Pacheco, L., H. Delgado, and V.E. Cabrera. 2021. Exploring machine learning algorithms for early prediction of clinical mastitis. International Dairy Journal 119:105051. doi:10.1016/j.idairyj.2021.105051.

Gengler, N. 2019. Symposium review: Challenges and opportunities for evaluating and using the genetic potential of dairy cattle in the new era of sensor data from automation. Journal of Dairy Science 102:5756–5763. doi:10.3168/jds.2018-15711.

Hogeveen, C., C. Kamphuis, W. Steeneveld, and H. Mollenhorst,. 2010. Sensors and Clinical Mastitis—The Quest for the Perfect Alert. Sensors 10:7991–8009.

Koltes, J.E., J.B. Cole, R. Clemmens, R.N. Dilger, L.M. Kramer, J.K. Lunney, M.E. McCue, S.D. McKay, R.G. Mateescu, B.M. Murdoch, R. Reuter, C.E. Rexroad, G.J.M. Rosa, N.V.L. Serão, S.N. White, M.J. Woodward-Greene, M. Worku, H. Zhang, and J.M. Reecy. 2019. A Vision for Development and Utilization of High-Throughput Phenotyping and Big Data Analytics in Livestock. Front. Genet. 10. doi:10.3389/fgene.2019.01197.

Lioutas, E.D., C. Charatsari, G. La Rocca, and M. De Rosa. 2019. Key questions on the use of big data in farming: An activity theory approach. NJAS - Wageningen Journal of Life Sciences 90–91:100297. doi:10.1016/j.njas.2019.04.003.

Lovarelli, D., J. Bacenetti, and M. Guarino. 2020. A review on dairy cattle farming: Is precision livestock farming the compromise for an environmental,

economic and social sustainable production?. Journal of Cleaner Production 262:121409. doi:10.1016/j.jclepro.2020.121409.

Menéndez González, S., A. Steiner, B. Gassner, and G. Regula. 2010. Antimicrobial use in Swiss dairy farms: Quantification and evaluation of data quality. Preventive Veterinary Medicine 95:50–63. doi:10.1016/j.prevetmed.2010.03.004.

Morota, G., R.V. Ventura, F.F. Silva, M. Koyama, and S.C. Fernando. 2018. Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. J. Anim. Sci. 96:1540–1550. doi:10.1093/jas/sky014.

Newton, J.E., R. Nettle, and J.E. Pryce. 2020. Farming smarter with big data: Insights from the case of Australia's national dairy herd milk recording scheme. Agricultural Systems 181:102811. doi:10.1016/j.agsy.2020.102811.

USDA. 2021. Data Science for Food and Agricultural Systems. Accessed April 14, 2021. https://nifa.usda.gov/resource/dsfas-faq#_Toc61252312.

Wolfert, S., L. Ge, C. Verdouw, and M.-J. Bogaardt. 2017. Big Data in Smart Farming – A review. Agricultural Systems 153:69–80. doi:10.1016/j.agsy.2017.01.023.