



17. Data Governance How do Organizations Deal with Stewardship of Farmer Data?

Title presentation

Collecting, Integrating, Harmonizing and Connecting Data from Dairy Farms: The US Dairy Brain Project experience

Author(s)

V.E. Cabrera, L. Fadul-Pacheco, S. Wangen, R. Fourdraine, & J. Mattison

Institution for which the first author of this abstract is working

University of Wisconsin-Madison, 1675 Observatory Dr., Madison, WI 53706, US.

Abstract

Modern dairy farms in the US generate vast amounts of data, and with constantly emerging new technologies and implementations, the frequency, diversity, and sheer quantity of this data is increasing. While each source of data can be valuable on its own, the integration of this data from different on-farm sources offers a worthwhile opportunity to add significant value to the processes of farm management and decision-making. While this diversity in data collection platforms can be beneficial to the farmer, and to the adoption of automation procedures in general, this diversity also complicates the integration of data from all of these different systems. There is currently no standardization of data organization from a single source (e.g., milking parlor data), meaning that successful integration of data at a large scale requires that each individual source x vendor combination has a unique standardization process to make the data interchangeable and generically available to analysis algorithms. As part of the University of Wisconsin Dairy Brain project, we are developing an Agricultural Data Hub (AgDH) to do just that. The AgDH is a system which obtains data from on-farm sources and implements a variety of parsing scripts, each one designed to handle the translation of data from one source x vendor combination into a source-specific (but vendor-generic) format. This standardized data structure is then stored and organized in a way that reflects the relationships and interdependencies between different on-farm data sources, facilitating the integration of on-farm data sources, and making the data available for further analysis. The AgDH is implemented using an extensible system of Directed Acyclic Graphs (DAGs) that orchestrate container-based standardization algorithms and Structured Query Language (SQL)-based data storage along with a data-serving Application Programming Interface (API) that makes it available to analytical services further down the value chain.