Available online at www.sciencedirect.com
ScienceDirect
Computers and Electronics in Agriculture 59 (2007) 39–55
Computers and electronics in agriculture
www.elsevier.com/locate/compag

# Data modeling for Precision Dairy Farming within the competitive field of operational and analytical tasks

Christian Schulze [a], Joachim Spilke [a,*], Wolfgang Lehner [b]

[a] *Institute of Agricultural and Nutritional Sciences, Martin-Luther-University Halle-Wittenberg, 06099 Halle, Germany*
[b] *Institute for System Architecture, Dresden University of Technology, 01062 Dresden, Germany*

Received 13 December 2006; received in revised form 10 May 2007; accepted 10 May 2007

**Abstract**

The interdisciplinary concept of Precision Dairy Farming sets very high standards for data management. Special consideration during its implementation must therefore be given to support both operational and analytical data uses (e.g., OLAP). The inclusion of both data views results in the data modeling being a hybrid of two conceptual design models. In contrast to previous design concepts, we will assume a parallel modeling process for both views, which results in a shared logical data schema. This is the only way to effectively avoid redundancies and inconsistencies on both the schema and data levels.

Using an ongoing application as an example, we will explain both methods and results. In doing so, we will make use of the Entity-Relationship Model (E/RM) for modeling operational data. We will also make use of E/RM's multi-dimensional extension, the multi-dimensional Entity-Relationship Model (mE/RM), for modeling analytical data. In order to meet all application-specific modeling requirements, however, new representation elements must be introduced. Therefore, we propose both a property window to describe the subject of analysis, and also a marker for temporal restrictions to the values of analysis structures as an extension of the mE/RM. Starting from the two conceptual models, we will then describe the logical modeling in a shared relational schema. Both the transformation of conceptual notation elements into relational structures and the creation of a required meta model will be explained during this step.

The procedures discussed in this paper are important for a variety of tasks in the field of Precision Dairy Farming and beyond.


*Keywords:* Data modeling; Precision Dairy Farming; Multi-dimensional Entity-Relationship Model; Relational model; OLAP; Data warehouse

## 1. Introduction

More than ever, the agriculture and food production industries are facing the challenge of establishing a permanent link between consumer protection, animal welfare, quality control and economic sustainability. The social status and acceptance of the farming industry is considerably dependent on how this challenge is met. With its complex requirements, tight interdisciplinary cooperation and a suitable systematic approach are needed. Such an approach is emerging with the concept of "Precision Livestock Farming". Moreover, its specific implementation for dairy cows, so-called "Precision Dairy Farming" (PDF), focuses on the transition from a group-oriented perspective to a perspective where intensive consideration is given to individual cows, and where particular goals, such as tapping individual potential, diagnosing diseases early, and using minimum medication are pursued (Spilke et al., 2003). Due

* Corresponding author. Tel.: +49 345 55 22690; fax: +49 345 55 27128.
*E-mail address:* joachim.spilke@landw.uni-halle.de (J. Spilke).

0168-1699/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.compag.2007.05.001

to its dominant competitive strength and immense importance for the agrarian sector, the dairy industry requires immediate action. Here, the transition to loose-housing systems – required both for animal welfare and economic reasons – and ever-growing herd sizes must not lead to a decrease in the intensity of care provided to individual cows. A variety of individual scientific research results and technological developments, especially within the field of data acquisition and process control, have laid a solid foundation for achieving the abovementioned goals. However, the next wave of innovation depends on whether or not it will be possible to integrate available data for diverse tasks and to permanently store them. In doing so, both operational short-term tasks and strategic long-term arrangements in the respective agrarian enterprises have to be supported; additional consideration should be given to documentation requirements of external institutions with regard to environmental protection, animal welfare and consumer protection, and even business processes with suppliers and clients should be included in the planning. When meeting these diverse requirements for the implementation of PDF, a central role will be assigned to the process of information management and, in particular, to data management.

The reality of our agricultural industry, however, is marked by considerable deficits. The use of relational databases for operational management programs is an exception rather than the rule (Bosch, 2001). This results in serious hindrances to further improvements in fields, such as application integration (Doluschitz, 2002) or the organization of production chains (Schiefer, 2002). Linseisen et al. (2000) and Linseisen (2001) show that powerful data management also plays a key role in the implementation of Precision Farming, a concept identical to PDF for the crop industry. Support for strategic decisions through analytical databases in the sense of data warehouses, as used and implemented intensively in the industrial sector (Kimball et al., 1998; Chamoni and Gluchowski, 1999; Mucksch and Behme, 2000), has thus far not been given serious consideration in the agrarian sector. Interestingly enough however, the importance of these processes, which applies to the implementation of PDF as well, has already been emphasized by Swensson and Sederblad (1997).

This paper attempts to remedy the abovementioned deficits. In order to do so, a certain type of data modeling will be used, which meets the specific requirements of PDF. This means both operational and analytical needs must be taken into consideration. Previous modeling approaches have already assumed a strict separation of operational and analytical tasks on the modeling level (Kimball et al., 1998). This is based on the fact that analytical tasks often use pre-existing operational data records whose modeling stage was completed at a much earlier point in time. The disadvantage of this kind of procedure is the inability to consider analytical problems when modeling operational data. Among the problems arising from this are often inefficient or even impossible analytical evaluations. Since the data management process within the framework of PDF needs to be revised in any case, these types of problems can be avoided. Therefore, we will follow an approach that considers both the operational and analytical use of data during the modeling stage. The goal of this paper is to present the design results for operational data first, and to show the extension of an analytical schema afterwards. Thus, we will make use of the Entity-Relationship Model (E/RM) and its multi-dimensional extension (mE/RM). A partial model of PDF will be used to point out the problems and deficits of these modeling approaches, to illustrate necessary extensions, and to implement them in a practical manner. In order to compare agrarian enterprises, the use of such model extensions and modeling results will be explained through the ongoing application of the German Cattle Breeders Federation (Arbeitsgemeinschaft Deutscher Rinderzüchter, ADR) for the comparison of agrarian enterprises.

## 2. Area description and methods

### *2.1. Description of the application scenario*

The current application of data modeling within the framework of PDF is a project of the German Cattle Breeders Federation, (Arbeitsgemeinschaft Deutscher Rinderzüchter, ADR) and is being carried out to compare agrarian enterprises. Based on detailed production data from participating enterprises, aggregated data shall be provided for anonymous enterprise comparisons. For clarity, we will limit the example to the mapping of data on disease treatments (Fig. 1).

Within enterprises, cow-specific data on diseases treated and on production level, etc., are available from operational systems. Both analysis tasks within the enterprises and comparisons between enterprises, however, require an initial aggregation to cow-group-related data based on a given period of time as a first step in the enterprise itself. Consequently, the resulting data cube for mapping the frequency of disease treatments is characterized by the disease, the time period,